جامعة الزقازيق
كلية التجارة

مقدمة الإحصاء

دكتور / حسن محمد علي
أستاذ م. بقسم الإحصاء والرياضة والتأمين

مطبعة الفردوس ت: ٠٥٥/٢٣٠٩٠٩٠

بسم الله الرحمن الرحيم

﴿لِيَعْلَمَ أَنْ قَدْ أَبْلَغُوا رِسَالَاتِ رَبِّهِمْ وَأَحَاطَ بِمَا لَدَيْهِمْ وَأَحْصَى كُلَّ شَيْءٍ عَدَدًا﴾

صدق الله العظيم

سورة الجن: الآية (٢٨)

تقديم

الحمد لله الذى هدانا لهذا وما كنا لنهتدى لولا أن هدانا الله ...

مما لا شك فيه أن علم الإحصاء يلعب دورا هاما فى حياتنا اليومية بحيث أصبح يشغل مكانا مرموقا بين بقية العلوم الأخرى، ومع تزايد إعتماد الباحثين فى كافة الميادين على علم الإحصاء واتساع نطاق استخدامه فى مختلف المجـــالات. فقد تعددت مراجعه وازداد الاهتمام بمبادئه وطرقه وأساليبه.

فعلى سبيل المثال، ففى مجال الدراسات والبحوث التجارية والاقتصاديـــة تظهر الحاجة إلى الإحصاء مرتبطة برغبة الدارس والباحث فى هذه المجالات إلى جمع وتلخيص البيانات المتصلة بموضوعات اهتمامه فى محاولة إلـــى الوصـــول إلى فهم مبدئى لأبعاد تلك الموضوعات. ما يؤثر عليها وما يتأثر بها، واســـتقراء لما تتضمنه البيانات من اتجاهات وعلاقات وينتهى به الأمر إلى الوصـــول إلـــى تفسيرات تحمل فى طياتها إجابات لتساؤلات تلح فى ذهنه لتشمل تعميمات تخـــص الحاضر وتنبؤات تمتد إلى المستقبل.

ومما يسعدنى أن أقدم هذا المؤلف كمقدمة فى علم الإحصـــاء الوصفـــى والذى يتضمن أساسيات الإحصاء الوصفى ومبادئه وتطبيقاته التى تمثل بدورهـــا حجر الأساس ونقطة الإنطلاق للإستقراء الإحصائى المتقدم بصفة عامة.

ويتناول هذا الكتاب عرض للعديد من الطرق والأساليب الإحصائية التـــى تساعد الباحثين والمهتمين بتطبيق علم الإحصاء فى كافة المجالات سواء أكـــانت اقتصادية أو تجارية أو اجتماعية أو غيرها ... وعموما فإن هذا المؤلف ما هو إلا مقدمة لابد من معرفتها قبل دراسة الإحصاء التحليلى والإحصاء التطبيقى، حيـــث يساعد هذا المؤلف على إمداد الباحث ببعض الطرق والأساليب الإحصائية التـــى تساعده فى تلخيص التجربة محل الدراسة فى صورة مجموعـــة مـــن المقـــاييس

والمؤشرات العلمية الضرورية لتحقيق هدف المشكلة محل الدراسة. لهذا فقد احتوى المؤلف على طرق جمع البيانات وتبويبها وعرضها بيانياً وجدولياً مع عرض مختصر لطرق المعاينة ثم الانتقال إلى عملية التحليل والتى تضمنت موضوعات فى مقاييس النزعة المركزية والتشتت والإلتواء والتفرطح فى التوزيعات التكرارية ... وهى مقاييس تهتم بدراسة متغير واحد ووصفه وتحديد طبيعته وكذلك كيفية توزيع مشاهداته، بعد ذلك ثم الإنتقال إلى بحث العلاقة بين متغيرين وإلى كيفية الإستفادة منها وطرق قياسها. وكذلك تضمن المؤلف قدراً معقولا من تحليل الإنحدار الخطى البسيط وبعض تطبيقاته.

وبعد - فأرجو من الله العلى القدير أن أكون قد وفقت فى تناول وعرض الموضوعات التى تضمنها هذا المؤلف، ولا أقول فى اختيار الموضوعات، فالموضوعات - أولاً وأخيراً - يحكمها المستوى الذى يعد المؤلف من أجله والوقت المتاح لدراسته.

ونسأل الله أن يلهمنا الصواب ويجنبنا الخطأ وعلى الله قصد السبيل. والله ولى التوفيق.

المؤلف

د/ حسين محمد على

يناير ٢٠٠٣م

الإهــداء

إلى روح أبى الطاهــرة

إلى أولادى

منى

أحمــد

دينـــا

نـــوران

أهدى هذا الكتاب

د/ حسن محمد على

فهرس الكتاب

تـقديـــم	١
الباب الأول : مفهوم علم الاحصاء ومراحل البحث الاحصائى	٣
١-١ مقدمه	٣
١-٢ مراحل البحث الاحصائى	٨
١-٣ أنواع البيانات الاحصائية	١٢
الباب الثانى : التوزيعات التكرارية وتمثيلها بيانيا	٣٣
٢-١ أساليب القياس وأنواع البيانات	٣٤
٢-٢ العرض الجدولى للبيانات	٣٨
٢-٣ العرض البيانى للبيانات	٦٦
الباب الثالث : النزعة المركزية	٩٣
٣-١ الوسط الحسابى	٩٤
٣-١-١ حساب الوسط الحسابى من البيانات غير المبوبة	٩٥
٣-١-٢ حساب الوسط الحسابى من البيانات المبوبة	١٠٧
٣-٢ الوسيط	١١٧
٣-٢-١ حساب الوسيط من بيانات غير مبوبة	١١٨

٣-٢-١ حساب الوسيط من بيانات غير مبوبة ١١٨

٣-٢-٢ حساب الوسيط من بيانات مبوبة ١٢٠

٣-٢-٣ شبيهات الوسيط الربيعين ١٢٧

٣-٣ المنوال ١٣٤

٣-٣-١ حساب المنوال فى حالة البيانات غير المبوبة ١٣٤

٣-٣-٢ حساب المنوال فى حالة البيانات المبوبة ١٣٥

٣-٤ الوسط الهندسى ١٤٤

٣-٤-١ حسابه فى حالة البيانات غير المبوبة ١٤٤

٣-٤-٢ حسابه فى حالة البيانات المبوبة ١٤٦

٣-٥ الوسط التوافقى ١٤٩

٣-٥-١ حسابه فى حالة البيانات غير المبوبة ١٤٩

٣-٥-٢ حسابه فى حالة البيانات المبوبه ١٥٠

الباب الرابع : مقاييس التشتت ١٦٥

٤-١ مقاييس التشتت المطلقة ١٦٩

٤-١-١ المدى ١٦٩

٤-١-٢ نصف المدى الربيعى ١٧٣

٤-١-٣ الانحراف المتوسط ١٧٥

٤-١-٤ التباين والانحراف المعيارى ١٨٢

٤-٢ مقاييس الاختلاف النسبى والقيم المعياريه ٢٠٦

٤-٢-١ مقاييس الاختلاف النسبى ٢٠٦

الباب الخامس : العزوم والالتواء والتفرطح ٢٢٥

٥-١ العزوم ٢٢٦

٥-١-١ العلاقة بين الزوم الصفرية والعزوم المركزية ٢٣٧

٥-٢ الالتواء ٢٤١

٥-٣ التفرطح ٢٤٥

الباب السادس : الارتباط الخطى ٢٤٩

٦-١ معامل ارتباط بيرسون ٢٥٥

٦-٢ ارتباط الرتب ٢٧٠

٦-٢-١ معامل اسبيرمان لارتباط الرتب ٢٧١

٦-٢-٢ معامل كندال لارتباط الرتب ٢٧٥

٦-٣ الارتباط بين الصفات ٢٧٩

٦-٣-١ معامل الاقتران ٢٨٠

٦-٣-٢ معامل التوافق ٢٨٢

الباب السابع : الانحدار الخطى [illegible]

٧-١ الانحدار الخطى بين المتغيرين (س،ص) فى حالــــة البيانات غير المبوبة ٢٨٦

٧-١-١ خط انحدار ص/س ٢٨٧

٧-١-٢ خط انحدار س/ص ٢٩٠

٧-٢ كيفية استخدام الانحدار الخطى البســـيط فى تعيين الاتجاه العام للسلاسل الزمنية ٣٠٧

الباب الأول
مفهوم علم الاحصاء ومراحل البحث الاحصائي
Statistics & Stages of Statistical Research

(١-١) مقدمـة :-

ظهرت أهمية علم الاحصاء منذ زمن بعيد نسبيا ، حيث يلعب هذا العلم دورا أساسيا وهاما في خدمة المجـالات العلميـة والعمليـة الأخـرى . إذ تدخـل الاحصاء بصورة أو بأخرى في حياتنا اليومية كأفراد وجماعات وكهيئات ودول دون أن نقصد أن نكون إحصائيين ، فنا نستخدمه كعلم وفن في البحث والدراسة والتحليل لمختلف الظواهر والمتغيرات بقصد الوصول الى مفهوم صحيح لهذه المتغيرات واتخاذ قرارات ما كان يسهل الوصول اليها الا باستخدام الاسـلوب الاحصائي .

ولقد ظهرت تعريفات عديدة لكلمة إحصاء ، يضع كل منها حدودا لما يمكن ان يحتويه علم الاحصاء ، وقد لا يكون من اللازم في هذا المجال أن نناقش الاختلافات اللفظية بين التعريفات المختلفة للاحصاء ولكن يجدر بنا أن نبين الحدود التي تضعها بعض هذه التعريفات .

فقد عرف علم الاحصاء قديما بأنه علم العد (*Sience of Counting*) ثم جمع هذه البيانات وعرضها في شكل جداول ورسوم بيانية تمهيدا لحساب

المؤشرات والمقاييس مثل مقاييس النزعة المركزية ومقاييس التشتت . وظل هذا المفهوم سائدا فترة طويلة عندما كان الهدف الرئيسى لهذا هو جمع البيانات عن عدد السكان ووضعهم بالمناطق المختلفة وأيضا حصر الممتلكات بقصد تقدير الضرائب التى يمكن جبايتها من السكان والتعرف على قدرات الدولة المادية والبشرية .

ثم تطور مفهوم علم الاحصاء إلى المفهوم الذى يدل على أنه علم التعامل مع المجتمعات المختلفة والتى ترجع الى قانون الاعداد الكبيرة ذات الاسباب العشوائية وعوامل الصدفة ، والمجتمع هنا لا يعنى بالضرورة المجتمع البشرى بل يعنى أى مجتمع لأفراد أو أشياء أو أدوات أو مفردات تجمع بينها صفة أو صفات مشتركة . فقد تتناول الدراسة الاحصائية مجموعة من البشر وقد نتناول بالدراسة والتحليل أيضا مجموعة نتائج لتجربة تتركب تلقائيا أو تجريبيا فى أى مجال من المجالات المختلفة .

والأمثلة على ذلك عديدة ويصعب حصرها . فدراسة السوق من أجل إنشاء مشروع جديد أو تطوير مشروع قائم وهو ما يسمى بدراسة الجدوى (*Feasiblity Study*)أو من أجل تسويق (*Marketing*) سلعة ما . وكذلك الرقابة على منتج ما للتحقق من مدى مطابقته للمواصفات وهو الاسلوب المعروف بضبط الانتاج (*Quality Control*) ، وتحديد معايير تمكن محاسب التكاليف من تحديد التجاوزات فى تكلفة الانتاج ، وخبير الاعلام الذى يود أن

يقيس أثر حملة اعلانية على تسويق فكرة ما أو برنامج معين مثل برنامج لتنظيم الاسرة ، وخبير التعليم الذى يرغب فى دراسة فعالية طريقة جديدة فى التدريس أو حين يود أن يخطط للمستقبل ليعد لهم المدرس والكتاب والمرافق التعليمية اللازمة والخبير الصحى حين يقيس أثر التلوث البيئى بسبب وجود عوادم ومخلفات صناعية وأثر ذلك على صحة المواطنين المقيمين فى المنطقة وكذلك على العاملين فى تلك الصناعة بهدف تنفيذ برامج صحية وطبية لتخفيض درجة التلوث من جهة ومعالجة الاثار الناجمة عنه من جهة أخرى ، وكذلك حين يقوم باحث باختبار فاعلية مصل جديد ، والباحث الزراعى الذى يقوم بدراسة تأثير نوع مستحدث من البذور أو السماد أو استخدام اساليب زراعية حديثة على انتاجية محاصيل زراعية مختلفة ،والتخطيط العمرانى للمناطق الصحراوية والمدن الجديدة ووضع سياسة رشيدة للسياحة والتنمية الاقتصادية - الاجتماعية على مستوى المشروع والدولة ..... الخ .

كل هذا لا يتحقق إلا على أساس رقمى يشمل تحديد وجمع البيانات ثم عرضها فى صورة يسهل معها استيعاب موضوع الدراسة ويلى ذلك تحليل البيانات بقصد الوصول الى قرارات محدده . وباختصار وباستخدام الطريقة الاحصائية فى دراسة الموضوع الذى يعنى الباحث أيا كان مجاله واهتماماته نجد أن هذه الطريقة تساعده على فهم واستيعاب ما حدث فى الماضى لاستخدامه فى التنبؤ بالمستقبل ورسم الخطط المستقبلية اللازمة .

وقد أدى التطور الهائل فى أساليب الرياضيات البحتة والتطبيقية وبالأخص نظرية الاحتمالات وكذلك تطور الحاسبات الآلية بأنواعها المختلفة وقدرتها الهائلة وطاقتها المتلاحقة إلى تطور علم الاحصاء وامكانية التطبيق العملى للنظريات والوصول عن طريقة الى أدق النتائج بسرعة فائقة .

مما سبق يمكن أن نعرف علم الاحصاء (*Statistics*) " بأنه العلم الذى يعنى بجمع وتنظيم وعرض وتحليل وتفسير البيانات الرقمية لمتغير أو أكثر فى مجتمع أو أكثر بغرض المساعدة فى الوصول الى قرار أفضل " ، وبايجاز فانه يمكن القول بأن الدراسة الرقمية لمتغير أو أكثر فى مجتمع أو عدة مجتمعات وأيا كان التعريف فانه يؤكد أن كل متغير يمكن قياسه رقميا سواء كان كميا أو وصفيا أو ترتيبيا ويمكن أن يخضع للدراسة الاحصائية ، كما أن دراسة هذا المتغير قد تكون على أســـاس الحصر الشامل (*Complete Enumeration*) لجميع المفردات فى المجتمع كله أو على أساس جزء يقتطع بصورة أو بأخرى من ذلك المجتمع وهو ما يطلق عليه أسلوب العينة (*Sample*) .

* الاحصاء الوصفى والاحصاء الاستنتاجى (التحليلى) :-

من التعريف السابق لعلم الاحصاء يمكننا أن نميز بين نوعين رئيسيين من الاساليب أو الطرق الاحصـــائية حسب الوظائف التى يؤديها كل منهما ، هما الاحصاء الوصـفى والاحصـاء الاسـتنتاجى وليس معنى ذلك التميز أنهما علمان منفصـــلان ولكنهما مكملان بحيث يكونان معا علم الاحصاء ، فالاحصاء الوصفى

(Descriptive Statistics) يهتم بتصنيف وترتيب البيانات التى تم جمعها عن الظاهرة أو الظواهر موضع الدراسة ووضعها فى جداول مختصرة أكثر سهولة فى الاستخدام مع تزويدها إن أمكن برسوم ايضاحية مناسبة وكذلك تلخيص هذه البيانات عن طريق حساب مقاييس وصفية لها دلاله واضحة كمقاييس المتوسط والتشتت بالاضافة الى مقاييس للعلاقة بين المتغيرات المختلفة التى تتعلق بهذه الظواهر وذلك لاظهار خصائص واتجاهات الظاهرة موضع الدراسة .

أما الاحصاء الاستنتاجى أو التحليلى (*Inferential Statistics*) فنقصد بها التوصل الى استنتاجات عامه عن مجموع المفردات التى يتشكل منها مجتمع الظاهرة موضع الدراسة وتسمى بالمجتمع (*Population*) وذلك من خلال دراسة مجموعة صغيرة من هذه المفردات تسمى بالعينة (*Sample*) ، أى التقييم من الجزء الى الكل ، وذلك لأن القيم الحقيقية المجتمع عادة ما يتعذر معرفتها ، وهو ما يفيد فى التخطيط والتنبؤ بما ستكون عليه الظاهرة محل الدراسة فى المستقبل .

والهدف من هذا المؤلف هو التعرض بشىء من التفصيل لأهم موضوعات وجوانب الشق الأول (الاحصاء الوصفى) أما الاحصاء الاستنتاجى فسيعرض ان شاء الله تعالى فى مؤلف آخر .

مما لا شك فيه أن حل المشاكل المختلفة التى تعترضنا سواء على المستوى القومى أو المستوى الفردى ، وفى ظل مواردنا المحدودة ، يتطلب بالضرورة استخدام الاسلوب المنظم تجاه ما يحيط بنا من ظواهر ومظاهر ومحاولة حل ما يعترضنا من مشاكل بأقل تكاليف ممكنة .

والبحث الاحصائى ، شأنه فى ذلك شأن أى بحث علمى ، يمر بالمراحل الاربعة الاتية :

أولا : تحديد المشكلة ووضع الفروض .

ثانيا : جمع البيانات .

ثالثا : عرض البيانات وتبويبها .

رابعا : تحليل البيانات وتفسيرها .

وفيما يلى استعراض موجز لمفهوم كل مرحلة من المراحل الاربعة السابقة :

<u>أولا : تحديد المشكلة ووضع الفروض</u>

عند التعرض لمشكلة ما يجب على الباحث حينئذ تحديد هذه المشكلة تحديدا دقيقا على أن يكون ملما الماما كافيا بجميع جوانبها ، فاغفال هذه الخطوة أو جانبا منها سوف يترتب عليه بالقطع التوصل الى نتائج غير دقيقة وقد تكون خاطئة . فيجب على الباحث أن يصيغ المشكلة موضع البحث فى كلمات ومعانى

محددة ودقيقة تقبل التحليل والدراسة العلمية وأن يبتعد بقدر الامكان عن العموميات والمعانى التى تتحمل أكثر من تفسير ، وأن يحدد من البداية الاهداف المرجو تحقيقها من وراء حل هذه المشكلة ولتحقيق هذا الغرض فمن الافضل وضع المشكلة فى صيغة مجموعة من الفروض الممكنه اختبارها فالفرض هو تفسير مبدئى للظاهرة موضع الدراسة ، وهذا التفسير يحتاج الى بيانات يتم جمعها وتحليلها ثم يقرر الباحث فى ضوء ذلك اما قبول الفرض كليا أو جزئيا أو رفضه والبحث عن فرض بديل .

ومن الواضح أن تحديد الفروض التى تعتقد أنها تؤثر فى المشكلة موضع البحث سوف تحدد للباحث منذ البداية البيانات الواجب جمعها ، وبالتالى فلا ينشغل بجمع بيانات ليست لها صله بالمشكلة موضع البحث فيتوفر الوقت والجهد والمال المقترح لاتمام البحث .

<u>ثانيا : جمع البيانات .</u>

وهو تجميع البيانات عن الظاهرة موضع الدراسة بطريقة علمية ، وجمع البيانات لا يمثل غاية ما نسعى اليه بل هى وسيلة نبغى من ورائها وصف الظواهر لدراسة طريقة تغيرها ومقارنتها بظواهر أخرى بغية استنباط العلاقات التى تربط بينها ولأهمية جمع البيانات اشترطت الحكومات أن يتم الحصول منها على اذن مسبق قبل البدء فى عملية جمع البيانات لدواعى كثيرة منها ما يتعلق بالامن القومى ومنها ما يتعلق بالمحافظة على قدرة الدولة فى منافسة الدول الأخرى .

ثالثا : عرض البيانات وتبويبها

بعد الحصول على البيانات الاحصائية وتجميعها سواء باستخدام أسلوب الحصر الشامل أو أسلوب العينة يكون لدينا فى الغالب عددا ضخما من الارقام والبيانات الاحصائية ، وهذه البيانات فى صورتها الاولية أو الخام يصعب استيعابها أو الاستفادة منها ، فمهما أوتينا من الدقة وحسن التتبع فلن يقدم لنا استعراض وتأمل هذه البيانات بطريقة مباشرة الا القليل جدا عن مدلول هذه البيانات وتفسيريها وتغيرها بالنسبة لبعضها البعض ، ومدى هذا التغير . لذلك فمن الضرورى ترتيب وتلخيص البيانات فى صورة تمكن الباحث من استقراء بعض العلاقات والاستنتاجات الاولية قبل البدء فى عملية التحليل الاحصائى وهى الغاية النهائية لاى دراسة . وعادة ما يتم اجـــراء عملية تبويب البيانات وعرضهـــا باحدى الطرق الاتية :

أ – تبويب البيانات وعرضها فى صورة جداول تكرارية .

ب – تبويب البيانات وعرضها فى صورة خرائط أو رسوم بيانية .

وسوف نتناول هاتين الطريقتين لتبويب البيانات وعرضها بالتفصيل فى الباب الثانى .

رابعا : تحليل البيانات وتفسيرها .

بعد الانتهاء من جمع وتبويب وعرض البيانات الاحصائية فلابد من استخدام الاساليب الاحصائية لتحليل وتفسير البيانات واستخلاص النتائج عن الظاهرة محل الدراسة والذى يعتبر الهدف الرئيسى من الدراسة أصلا .

وتحليل البيانات يتم بحساب بعض المقاييس الاحصائية مثل مقاييس النزعة المركزية والتشتت والعزوم والالتواء والتفرطح والارتباط والانحدار وغيرها من المقاييس التى تمكن من معرفة طريقة توزيع الظاهرة محل الدراسة وتفسير التغيرات المختلفة للظاهرة وقياس مقدار هذه التغيرات مما يفيد فى التنبؤ بالقيم المستقبلية للمتغيرات للتعرف على مسار الظاهرة فى المستقبل .

كما أن تحليل البيانات يشمل ايضا اجراء بعض الاختبارات الاحصائية للوصول الى قرار بقبول أو رفض الفروض التى افترضت تفسيرات مؤقتة للظاهرة محل الدراسة وكذلك تحديد فترات الثقة للتقديرات التى تحسب من البيانات .

وسوف نعالج هذه الموضوعات بشئ من التفصيل فى الابواب التالية من هذا المؤلف .

(١-٣) أنواع البيانات الاحصائية :-

١- البيانات الكمية والوصفية :-

يمكن تصنيف البيانات الاحصائية إلى بيانات كمية وأخرى وصفية ، فالبيانات الوصفية هى تلك التى تمثل قياسات أو أعداد . والأمثلة على ذلك عديدة فمثلا . البيانات التى توضح أطوال مجموعة من الأشخاص أو تلك التى تبين أعداد العملاء المترددين على أحد المتاجر فى يوم ما .

أما البيانات الوصفية فهى تلك التى تمثل خصائص أو صفات لا يمكن قياسها ... كن فقط تصنيفها أو وصفها ومن أمثلة البيانات الوصفية تلك البيانات التى ... من حيث تصنيفه (أعزب أم أرمل ... الخ) .

٢- البيانات الاولية (الميدانية) والبيانات الثانوية (ال...

البيانات الاولية هى تلك البيـ..ت التى يقوم ال... حث بجمعها بنفسه (أو عن طريق من ينوب عنه) لتحقيق الغرض من جمع هذه البيانات . وعادة ما يقال فى مثل هذه الاحوال أن الباحث قد حصل على بياناته من المصادر الاولية أو الميدانية . فعلى سبيل المثال اذا أراد باحث دراسة الوضع الاقتصادى لسكان احدى المناطق فعليه القيام باجراء المقابلات الشخصية لسكان هذه المنطقة أو

ارسال استمارات تحتوى على مجموعة الاسئلة التى يريد الاجابة عليها ويطلق على البيانات التى تم تجميعها فى هذه الحالة بالبيانات الاولية أو الميدانية .

ولهذه الطريقة من طرق جمع البيانات بعض المزايا ، أهمها أنها تمكن الباحث من التخطيط الجيد والذى يساعده فى الحصول على البيانات المطلوبة بدقة عالية ، علاوة على تعرفه على أوجه القصور التى تقابله فى جمع البيانات ومحاولته السريعة لحل مثل هذه المشكلات .

أما البيانات الثانوية فهى بيانات لم تجمع بمعرفة الباحث شخصيا وانما جمعها غيره لأغراض أخرى . ومن أمثلة ذلك الاحصائيات التى تظهر فى الكتب أو الدوريات أو النشرات الاحصائية التى تصدرها المؤسسات العامة والخاصة (جهاز التعبئة العامة والاحصاء ، بعض المنظمات التابعة لهيئة الامم المتحدة .... الخ) .

فاذا أخذ باحث بياناته عن تلك المصادر فنقول انه أخذ بياناته عن مصادر ثانوية (أو تاريخية) . فمثلا اذا أراد باحث دراسة العلاقة بين حوادث المرور وأعداد السيارات وأخذ البيانات المتعلقة بالظاهرتين من سجلات ادراة المرور . حينئذ يقال أن الباحث قد استخدام بيانات ثانوية .

وتعتبر لهذه الطريقة من طرق جمع البيانات بعض العيوب ، متمثلة فى عدم إلمام مستخدم هذه البيانات بكافة الظروف المحيطة والكيفية التى جمعت بها هذه البيانات ، علاوة على حاجة هذه البيانات الى التعديل قبل استخدامها . لذا

يجب استخدامها بحذر شديد . الا أن هذه الطريقة لهـا بعض المزايا والتى تتمثل فى توفير الوقت والجهد فى جمع البيانات المطلوبة .

*** المتغير والثابت :-**

المتغير هو خاصية (مثل الطول ، الوزن ، المبيعات ، المشتريات ، العمر . . . . . الخ) حيث يأخذ عادة قيما مختلفة لمفرداته .

وعادة ما يتـم تقسيم المتغيرات الـى متغيرات منفصلـة أو متصلـة ، **فالمتغير المنفصل** هو الذى يأخذ عددا محدودا أو غير محدود من القيم (المنفصلة) . والامثلة على المتغير المنفصل عديدة مثل المتغير الذى يمثل عدد الاطفال الذكور فى الاسر داخل احدى القرى ، عدد الوحدات المنتجة داخل احد [illegible] . . . . الخ [illegible] المنفصل يأخذ قيم صحيحة سواء بصورة محدودة أو غير محدودة .

أما **المتغير المتصل** فهو المتغير الذى يمكن أن يأخذ أى قيمة (صحيحة أو غير صحيحة) فى نطاق تغيره . و أدل مثال على ذلك المتغير الذى يمثل أعمار مجموعة من الطلاب داخل فصل دراسى . أما الثابت فهو تلك المفردة التى تأخذ قيمة واحدة فقط لا تتغير .

*** أساليب الدراسة الميدانية :-**

من المعروف وجود أسلوبان للدراسات الميدانية (السابق التعرض لها) هى الحصر الشامل والعينات ولكل منهما مزاياه وعيوبه ، كما أن هناك بعض الحالات التى يكون فيها الباحث مضطر الى استخدام أسلوب دون الأخر بغض النظر عن المزايا والعيوب التى تصاحب كل سلوك .

أولا : أسلوب الحصر الشامل :

الدراسة عن طريق الحصر الشامل تعنى أن الدراسة (أو جمع البيانات) تشمل كل مفردات المجتمع الاحصائى . وعلى الباحث أن يحدد تحديدا دقيقا وواضحا ما هو المجتمع الاحصائى موضع الدراسة .

ويقصد بالمجتمع الاحصائى كل المفردات التى يجمعها إطار واحد وعام أو مجموعة خصائص عامة واحدة . والمجتمع الاحصائى بهذا المفهوم ليس قاصرا على المجتمع البشرى فحسب بل يتعدى ذلك الى ما هو أشمل وأعم ، فالمجتمع الاحصائى – على سبيل المثال ، قد يكون طلاب الفرقة الثالثة بكلية التجارة- جامعة الزقازيق فى فصل دراسى معين ، وقد يكون المجتمع الاحصائى متمثل فى اعداد السيارات المنتجة فى مصنع ما فى أحد السنوات ..... وهكذا . كذلك قد يكون المجتمع الاحصائى محدودا أى يمكن تحديد حجمه أو حصر جميع مفرداته ، كما قد يكون غير محدود أى لا يمكن معرفة حجم أو عدد مفرداته .

فالدراسة عن طريق الحصر الشامل تعنى أن الباحث سيقوم بجمع بيانات عن كل مفردات المجتمع الاحصائى بعد تحديده تحديدا دقيقا وواضحا .

*** مزايا أسلوب الحصر الشامل :-**

لاسلوب الحصر الشامل عدة مزايا أهمها :

١. أنه الاسلوب الاكثر ملائمة فى بعض الحالات مثل :-

- حالات التعدادات : مثل اجراء التعداد العام للسكان . تعداد القوى العاملة ، تعددات المنشأت الصناعية ..... الخ .
- الحالات التى قد يترتب على عدم دراستها أو فحصها بعض الاضرار : مثل التطعيم ضد مرض معين ، فحص أنابيب الغاز المنتجة بواسطة أحد مصانع تعبئة الغاز ... الخ.ففى مثل هذه الحالات يكون أسلوب الحصر الشامل هو الاسلوب الأكثر ملائمة فى جمع البيانات .

٢. أن استخدام أسلوب الحصر الشامل يتيح للباحث فرصة جمع بيانات عن جميع مفردات المجتمع موضع الدراسة ، وما يستتبع ذلك من الحصول على نتائج أكثر دقة وموضوعية .

*** عيوب استخدام اسلوب الحصر الشامل :-**

من الطبيعى انه كلما كان حجم المجتمع الاحصائى محل الدراسة كبيرا ومنتشرا على مساحات جغرافية واسعة كلما أدى ذلك الى ظهور عيوب استخدام الحصر الشامل والمتمثلة فى زيادة كل من الوقت والجهد والتكاليف لاجراء مثل هذه الاحصاءات .

ثانيا : أسلوب العينات :-

عادة ما يفضل استخدام أسلوب العينات فى جمع البيانات فى الحالات التى يكون فيها المجتمع الاحصائى لا نهائى وغير محدود ، وكذلك فى الحالات التى يكون مفردات المجتمع الاحصائى متجانسة وبصفة عامة يحبذ استخدام أسلوب المعاينة على وجه الخصوص بدلا من الحصر الشامل عدة اعتبارات تنحصر فى الوفر فى الوقت والجهد والمال وكذلك امكان الحصول على نتائج أكثر دقة اذا ما أحسن اختيار مفردات العينة بحيث تكون ممثلة للمجتمع الاحصائى موضع الدراسة تمثيلا جيدا .

ولقد شاع استخدام أسلوب العينات وعم كثيرا من ميادين الدراسات التجارية والاقتصادية والاجتماعية والصحية والعلمية . فقد أصبحت العينات وسيلة اساسية فى دراسة مشكلات التسويق كما تستخدم أيضا فى دراسة الهجرة والعمالة وميزانية الاسرة ودرجة شمول تعدادات السكان ودرجة جودة البيانات التى تجمع اثناءها كما استخدمت وتستخدم فى تقييم تسجيل الاحداث الحيوية من حيث درجة الشمول والدقة كما أصبحت العينات وسيلة المنتج لاختبار درجة انتاجه .

ولقد أخذت كثيرا من الدول والهيئات الدولية بهذا الاسلوب ومن أمثلة ذلك فى جمهورية مصر العربية بحث القوى العاملة (١٩٦١) وتقدير السكان بالعينة

(١٩٦٦) وبحث ميزانية الاسرة (١٩٧٤) والمسح القومى للخصوبة (١٩٨٠) وتقييم نتائج تعداد السكان لسنة (١٩٨٦) من حيث الشمول ودرجة الدقة .

*** مزايا أسلوب العينات :-**

يتميز أسلوب العينات بمزايا عديدة منها :

١-توفير الوقت والمجهود والتكاليف .

٢-انه الاسلوب الاكثر ملائمة فى الحالات التى يترتب على فحصها أو دراستها أهلاك أو افساد الوحدات محل الفحص أو الدراسة مثل فحص الدم ، فحص أو اختبار الاغذية المحفوظة (المعلبات) ، ..... الخ حيث يكون الاسلوب الاكثر ملائمة فى مثل هذه الحالات هو العينات .

*** عيوب أسلوب العينات :-**

على الرغم من المزايا العديدة لاسلوب العينات الا أن هذا الاسلوب يعتريه بعض العيوب منها :

١-لا يوفر للباحث بيانات عن كل مفردات المجتمع .

٢-تتوقف نتائج الدراسة الى حدا كبير على مدى تمثيل العينة للمجتمع .

وبرغم هذه العيوب الا أن أسلوب العينات أصبح - كما ذكرنا - هو الاسلوب الاكثر شيوعا خصوصا بعد تقدم علم الاحصاء الاستدلالى .

* **المعاينة والعينة :-**

قد يختلط الامر علـى القـارىء فـى بعض الاحيـان عنـد تمـيزه الفرق بين العينة والمعاينة ، الامر الذى يستوجب علينا توضيح هذين المصطلحين .

حيث تعرف **المعاينة** (*Sampling*) بأنها الاسلوب الذى نستطيع عن طريقـه الحصـول علـى بيانـات معينــة عـن مجموعـة مــن المفـردات (مجتمــع) باستخدام جزء فقط من هذه المجموعة يسمى عينة .

وقد انتشرت فى الاونة الاخيرة ابحاث يطلق عليها المسـموح (*Survay*) تجـرى لتحقيق أغراض معينة كأن نتعرف علـى جـدوى برنـامج معين لتنظيـم الاسـرة أو لاستطلاع اراء الناخبين بالنسبة لمرشـح معيـن أو لمـدى تقبل المسـتهلك لقـرارت ـادية معينة وما الى ذلك . كل هذه المسموح يستخدم فيهـا أسـلوب العينـات .
لذا يجب التعرض بصـ ـ أوضح لتعريف العينة .

ا **العينـة** (*Sample*) فتعرف على أنه جـزء يختـار بطريقـة احتماليـة بأسلوب أو ـ أو بطريقة غير احتمالية وذلك بغرض دراسة خصـائص المجتمـع الذى تسحب منه العينة .

وبصفة عامة يجب أن تتوافر فى العينات المسـحوبة بعض الشـروط التـى من أهمها :

١-أن تكون العينة المسحوبة ممثلة للمجتمع أحسن تمثيل ممكن .

٢-التقديرات التى تحسب من العينة لخصائص المجتمع محل الدراسة يجب ان تكون دقيقة ويمكن قياس درجة دقتها وهذا ما يطلق عليها بدرجة المأمونية للعينة (*Reliability*) .

٣-يجب أنْ تكون تكاليف البحث باستخدام أسلوب العينة منخفضة اذا ما قورنت بتكاليفه باستخدام أسلوب الحصر الشامل ولا يقصد هنا بالتكاليف المادية فقط بل الوقت والمجهود المبذول ايضا كعناصر هامة من عناصر التكلفة .

*** المعاينة الاحتمالية والمعاينة غير الاحتمالية :-**

يقصد **بالمعاينة الاحتمالية** (*The Probability Sampling*) هى العينات التى يتم اختيار مفرداتها من بين وحدات المعاينة التى يتكون منها المجتمع باسلوب احتمالى يوفر لكل وحده من وحدات المعاينة احتمال الاختيار ضمن مفردات العينة المسحوبة وهذا الاحتمال ثابت ومحدد تنحصر قيمته بين الصفر والواحد الصحيح .

أما **المعاينة غير الاحتمالية** (*The non Probability Sampling*) فيقصد به الاعتماد على التقدير والحكم الشخصى محل الاحتمال فى اختيار مفرادت العينة . وعادة ما يرجح استخدام هذا الاسلوب لبعض الاعتبارات العلمية أو السهولة فى الاختيار ، فعلى سبيل المثال قد يلجأ أحد منتجى سلعة ما الى بعض المدن التى يتوقع أنها تمثل سوق هذه السلعة من واقع خبرته الشخصية .

مما سبق يتبين لنا أن هناك كثير من الدراسات التى يتحتم فيها استخدام اسلوب العينة فى جمع البيانات ، كما أن كثير من الدراسات يمكن إجراؤها باسلوب الحصر الشامل وأيضا بأسلوب العينة ويفضل فيها الاسلوب الاخير لاعتبارات مادية وفنية. وعند جمع البيانات يمكن أن نقع فى نوعين من الاخطاء.

١- خطأ المعاينة :-

والذى ينتج عن استخدام أسلوب العينة فى جمع البيانات ويرجع الى اختلاف النتائج المتحصل عليها من العينة عن نتائج المجتمع ومن البديهى أن هذا النوع من الاخطاء لا يصادفنا عند استخدام أسلوب الحصر الشامل ، ويتوقف هذا الخطأ على بعض العوامل منها :

- **حجم العينة المختارة**

فكلما زاد حجم العينة المختارة كلما قل خطا المعاينة والعكس صحيح ، بمعنى أن خطا المعاينة يتناسب عكسيا مع حجم العينة المختارة .

- **درجة التجانس داخل المجتمع**

فكلما كانت مفردات المجتمع متجانسة فيما بينها كلما كان خطأ المعاينة أقل ، والعكس صحيح كلما قلت درجة التجانس داخل المجتمع كلما زاد خطأ المعاينة ، بمعنى أن خطأ المعاينة يتناسب طرديا مع درجة

تجانس المجتمع (تباين المجتمع) حيث يقصد بتباين المجتمع (مدى تشتت أو ابتعاد مفرداته عن بعضها البعض) وهذا ما سوف نتعرض له بشىء من التفصيل عند دراسة مقاييس التشتت فى الاجزاء التالية .

• **طريقة اختيار مفردات العينة**

حيث تؤثر طريقة اختيار مفردات العينة أيضا على خطأ المعاينة سواء بالزيادة أو النقص .

٢- خطأ التحـيز :-

هذا النوع من الاخطاء يرجع الى القصور فى امكانيات البحث والى تحيزه وأيضا الى جامع البيانات نفسه ، كما أنه قد يظهر نتيجة قيام مستخدم البيانات بالتقريب فى البيانات المس[illegible] التى تجمع سواء باستخدام أسلوب الحصر الشامل أو المعاينة .

وبالرغم من أن أسلوب الحصر الشامل معرض لنوع واحد فقط من الاخطاء (أخطاء التحيز) وأن اسلوب العينة معرض لخطأي التحيز والمعاينة ، إلا أن أخطاء التحيز فى اسلوب الحصر الشامل قد يكون أكبر من مجموع خطأى التحيز والمعاينة باستخدام اسلوب العينة .

وسوف نتناول فيما يلى بايجاز شديد الطرق المختلفة لاختبار العينة الاحتمالية (طرق المعاينة) .

وهناك عدة طرق لاختيار العينات العشوائية (الاحتمالية) ، ولكننا سوف نكتفى بدراسة أكثر طرق المعاينة الاحتمالية استخداما وهى :

1- العينة العشوائية البسيطة :- Simple Random Sample

يعتبر أسلوب العينة العشوائية البسيطة أبسط وأدق طرق المعاينة الاحتمالية وأكثرها اعتمادا على العشوائية والصدفة ، وفيها يتم اعطاء كل مفردة من مفردات المجتمع نفس فرصة الظهور والمشاركة فى العينة المسحوبة ، فاذا أردنا مثلا اختيار عينة عشوائية بسيطة مكونة من عشرة أفراد من مجتمع مكون من مائة شخص ، فيمكن اختيار العينة العشوائية البسيطة بعدة طرق منها : أن نعطى لكل فرد فى المجتمع رقما مسلسلا نضعه على بطاقة فيكون لدينا مائة بطاقة متماثلة فى الحجم والوزن والملمس ثم يتم خلط هذه البطاقات خلطا جيدا ثم نقوم بسحب عشرة بطاقات بطريقة عشوائية فتكون الارقام الموجودة على كل منها هى الافراد العشرة المكونة للعينة المطلوبة ويسمى الاشخاص العشرة فى هذه الحالة بمفردات العينة ، هذا الاسلوب المستخدم فى اختيار مفردات العينة يسمى بالاسلوب العشوائى لذلك سمى هذا الاسلوب من أساليب المعاينة بالعينة العشوائية البسيطة .

واضح أن هذه الطريقة لاختيار مفردات العينة تحتاج الى الكثير من الوقت والجهد لذلك لا يسهل استخدامها وخاصة اذا كان حجم المجتمع كبيرا وتنتشر مفرداته فى نطاق واسع ، وهناك طرق أخرى لاختيار مفردات العينة

العشوائية منها أستخدام جداول الاعداد العشوائية واستخدام الحاسبات الالكترونية والتى تمكننا من سحب العينات المطلوبة بسرعة ودقة فائقتين .

وبغض النظر عن الطريقة المستخدمة فى اختيار المفردات ولتوضيح المفاهيم الاساسية للعينة العشوائية البسيطة نفترض أن لدينا مجتمع عدد مفرداته (حجمه ن) وكان عدد مفردات العينة المطلوب اختيارها (حجمه ن) هو فان :

- عدد العينات العشوائية البسيطة التى يمكن اختيارها من المجتمع =

$$^{ن}ق_{ن} = \frac{ن!}{(ن - ن)!\ ن!}$$

- احتمال اختيار أى من العينات المسحوبة = $\frac{1}{^{ن}ق_{ن}} = \frac{(ن - ن)!\ ن!}{ن!}$

- احتمال اختيار أى مفرده بالعينة البسيطة = $\frac{\text{حجم العينة}}{\text{حجم المجتمع}} = \frac{ن}{ن}$

مثـــــال (١) :-

بافتراض أن حجم المجتمع عبارة عن أربعة مفردات هى أ ، ب ، جـ ، ء فاذا أردنا سحب عينة مكونة من مفردتين فانه يمكن حساب عدد العينات الممكن سحبها على النحو التالى :

حيث أن :

حجم المجتمع (ن) = ٤ ، حجم العينة (نه) = ٢ فإن :

عدد العينات الممكن سحبها من هذا المجتمع = $^{٤}ق_{٢}$ = $\frac{٤!}{٢! \; ٢!}$ = ٦ عينات

وهى (أ ، ب) ، (أ ، جـ) ، (أ ، ء) ، (ب ، جـ) ، (ب ، ء) ، (جـ ، ء)

• احتمال اختيار أى من العينات المسحوبة = $\frac{١}{^{٤}ق_{٢}}$ = $\frac{١}{٦}$

• احتمال اختيار أى مفردة بالعينة العشوائية البسيطة = $\frac{نه}{ن}$ = $\frac{٢}{٤}$ = $\frac{١}{٢}$

وتتميز العينة العشوائية البسيطة بسهولة اختيارها واعتمادها على عنصر الصدفة مما يؤدى الى دقة النتائج المتحصل عليها الى حد كبير . ولكن يعاب عليها أنه اذا كان حجم المجتمع مكونا من طبقات غير متجانسة من حيث الظاهرة موضوع الدراسة فان العينة العشوائية البسيطة قد لا تضمن تمثيل كل طبقة من هذه الطبقات فى العينة بنفس نسبتها أو وزنها فى المجتمع الاصلى وبذلك تكون العينة غير ممثلة للمجتمع وتصبح بالتالى النتائج المتحصل عليها غير دقيقة ، فاذا كان المجتمع محل الدراسة مثلا عبارة عن مصنع مكون من عمال وموظفين بنسبة ٧ : ٣ واخترت عينة عشوائية بسيطة من هذا المصنع لدراسة أثر الحوافز التى يمنحها المصنع على زيادة الانتاج مثلا ، فالعينة العشوائية البسيطة فى هذه الحالة قد لا تضمن لنا أن يكون الفئتان ممثلتين فى العينة بنفس نسبة تواجدهما

بالمصنع وهى ٧٠ : ٣ مما قد يؤدى الى ظهور خطأى المعاينة والتحيز (التحيز لفئة على حساب الاخرى) .

ويعاب ايضا على العينة العشوائية البسيطة أنه قد يصعب أختيارها من مجتمع كبير تنتشر مفرداته انتشارا واسعا ، اذ أن اختيار مفردات العينة بدقة وامانه سوف يؤدى الى زيادة تكاليف البحث زيادة لا مبرر لها .

٢ - *__العينة العشوائية المنتظمة Systematic Random Sample :-__*

تستخدم هذه الطريقة اذا كانت مفردات المجتمع موضع الدراسة متجانسة ومرتبة فى اطار معين ، فمثلا فى حالة اختيار عينة من الاجور لبعض العاملين من بين كشوف للاجور مرتبة حسب الدرجة المالية . وطريقة اختيار مفردات العينة العشوائية المنتظمة يتم على النحو التالى :

اذا كان حجم المجتمع = ن ، وحجم العينة المطلوب سحبها = نـ

وطبقا لاسلوب المعاينة المنتظمة يتم تقسيم مفردات المجتمع الى ن من المجموعات المتساوية من حيث العدد بحيث تكون :

$$\text{طول كل مجموعة} = \text{ك} = \frac{\text{نـ}}{\text{ن}}$$

ثم يتم اختيار المفردة الاولى عشوائيا بين مفردات المجموعة الاولى وليكن ترتيبها فى هذه المجموعة هو (أ) .

وبالتالى فان بقية المفردات التى ستنضم الى العينة يتحدد فورا وهى التى ترتيبها أ ، أ + ك ، أ + ٢ك ، . . . . ، أ + (ن-١) ك .

ومن الملاحظ أن عنصر العشوائية يتحقق فقط عند اختيار المفردة الاولى بينما عنصر الانتظام يتحقق عند اختيار باقى مفردات العينة ، وهذا هو السبب فى تسمية هذا النوع من العينات بالعينة العشوائية المنتظمة .

فمثلا ، اذا كان المجتمع يتكون من ١٠٠ مفردة وكان حجم العينة المطلوب اختيارها هو ن = ٥ ، فطبقا لهذا النوع من العينات يتم تقسيم مفردات المجتمع الى خمسة مجموعات متساوية الحجم بحيث أن :

$$\text{عدد المفردات داخل كل مجموعة} = \frac{١٠٠}{٥} = ٢٠ \text{ مفردة}$$

ثم يتم اختيار المفردة الاولى من المجموعة عشوائيا وليكن ترتيبها التاسع مثلا ، وبالتالى فان المفردات التى سوف تنضم الى العينة يكون ترتيبها هو : ٩ ، ٩ + ٢٠ ، ٩ + ٢(٢٠) ، ٩ + ٣(٢٠) ، ٩ + .(٢٠)

أى تكون مفردات العينة هى المفردات التى ترتيبها : ٩ ، ٢٩ ، ٤٩ ، ٦٩ ، ٨٩

وتجدر الاشارة الى أنه اذا كانت مفردات المجتمع المختارة فى العينة المنتظمة ذات ترتيب عشوائى فان نتائج العينة المنتظمة سوف تتفق مع نتائج العينة العشوائية البسيطة .

٣- العينة العشوائية الطبقية Stratified Random Sample :-

علمنا من مناقشتنا لاساليب المعاينة سواء العشوائية البسيطة أو المنتظمة أنه اذا كانت مفردات المجتمع متجانسة ، فإننا نفضل اختيار عينة عشوائية بسيطة لتمثيل ذلك المجتمع ، الا أنه اذا كان مجتمع الدراسة غير متماثل واستخدمنا أسلوب العينة العشوائية البسيطة فى اختيار مفردات العينة المطلوبة ، فان ذلك سيؤدى الى الحصول على عينة لا تمثل المجتمع تمثيلا جيدا .

لذلك وفى مثل هذه الحالات (عدم تجانس مفردات المجتمع) يفضل استخدام اسلوب المعاينة الطبقية . ويتم ذلك على عدة مراحل هى :

١. تقسيم مجتمع الدراسة الى مجموعة من الطبقات المتجانسة فيما بينها ، ولا يشترط فى هذه الحالة تساوى عدد المفردات داخل كل طبقة (كما هو الحال فى العينة المنتظمة) .

٢. تحديد الحجم المناسب للعينة التى سوف يتم اختيارها .

٣. تحديد عدد المفردات التى يجب اختيارها من كل طبقة .

٤. اختيار عينة عشوائية بسيطة من كل طبقة بنفس العدد الذى سبق تحديد اختياره من كل طبقة .

٥. تتكون العينة العشوائية الطبقية من مجموعة العينات العشوائية البسيطة المختاره من الطبقات المختلفة .

وقد تقسم الطبقات على أساس جغرافى كأن تقسم مدينة معينه الى مناطق جغرافية أو على أساس نوعى كتقسيم المصانع الى مصانع ذات حجم كبير وأخرى ذات حجم متوسط أو صغير . وعادة ما يتم استخدام العينة الطبقية فى المجالات الاتية :

١. اذا كانت هناك بيانات ذات دقة محدده مطلوب معرفتها لأقسام فرعية خاصة من المجتمع .

٢. أيضا قد يكون استخدام العينة الطبقية مناسبا فى الاستخدامات الادارية أو السياسية .

٣. كذلك يناسب استخدام ال[illegible]ة الطبقية معاينة الدخل وقد نجد عند دراسة الدخول [illegible]تع [illegible] [illegible]خل مرتفع وأخرى يحصل أفرادها على دخول منخفضة .

[illegible] [illegible]تخدا[illegible] [illegible] [illegible]معاينة اذا تجمع لدينا قيما متطرفة ذات الظاهرة [illegible]ضوع الدراسة ، فى هذه الحالة عادة ما يتم جمع هذه المفردات الشاذه فى طبقة منفصلة .

الا أن استخدام أسلوب العينة الطبقية يتطلب فى المقابل امكانية تقسيم الأصل الى طبقات متجانسة وأن تختلف كل طبقة عن الاخرى من حيث توزيع الظاهرة محل الدراسة وأن يكون حجم كل طبقة معروفا بدقة ووضوح ، اذ أن حجم كل طبقة سيؤثر فى عدد المفردات الواجب اختيارها من كل طبقة .

٤- العينة العشوائية متعددة المراحل (العنقودية) Cluster Sample:-

تستخدم العينة العشوائية العنقودية متعددة المراحل (عندما يكون حجم المجتمع كبيرا وقد تكون وحدات المعاينة واسعة الانتشار فى مجتمع يمتد لمساحات أو مسافات كبيرة) مما يرفع من تكلفة أختيار وتنفيذ الدراسة بالعينة بأى من أساليب المعاينة السابقة قد يحدث أن لا يكون اطار المعاينة معدا مسبقا وأن يكون تكاليف تحديثه أو اعداده مرتفعه لذلك فقد نختار أسلوب أخر للمعاينة وذلك بتقسيم المجتمع الى أقسام فرعية أو تجمعات فرعية متقاربة تشمل كل منها عدد من وحدات المعاينة (التى سوف تسجل عنها القراءات) ثم نختار من هذه الاقسام أو التجمعات الفرعية عدد منها لجمع البيانات منها ويعرف هذا الاسلوب بالمعاينة العنقودية ذات المراحل الواحدة . وقد نختار أن يعد اطار لكل وحده من وحدات المرحلة الاولى - والتى يطلق عليها اسم الوحدات الاولية للعينة ، ومن كل اطار نختار عدد من الوحدات النهائية وهى التى ستسجل عنها المشاهدات ، وهذا هو أسلوب العينة ذات المرحلتين ، ويتم اختيار مفردات المرحلة الثانية من وحدات المرحلة الاولى اما باسلوب العينة العشوائية البسيطة أو المنتظمة وبالتالى يكون أسلوب المعاينة العنقودية خليط من العينة العنقودية والعشوائية البسيطة أو المنتظمة بحسب الطريقة التى يتم بها اختيار وحدات المرحلة الثانية . وقد تتعدد المراحل ما بين اختيار الوحدات الاولية والوحدات النهائية حسب مقتضيات الدراسة وبذلك تكون العينة متعددة المراحل .

فمثلا اذا كانت الدراسة لمراجعة قسائم البيع فى شركة متعددة الفروع مثل شركة بيع المصنوعات المصرية وهى شركة ذات فروع عديدة واسعة الانتشار واذا كان حجم العينة هو ١٠ ٪ مثلا من قسائم البيع خلال سنة مالية ما أو فى شهر معين فان مراجعة ١٠٪ من القسائم فى جميع الفروع سوف يستغرق وقت ويستنفذ جهد كبير لذلك فقد يستعاض عن ذلك بالاقتصار على مراجعة قسائم البيع فى ١٠٪ من أنواع الشركة المذكورة وهو مثال لاسلوب العينة العنقودية ذات المرحلة الواحدة .

• وكمثال فعلى فقد اجرى تعداد السكان (١٩٦٦) فى ج . م . ع للعينة ذات السكانية – وفى إختيار عينة أسر الريف فقد اختيرت عينة طبقية عنقودية حجمها ٣٪ من القرى ، وقسمت كل محافظة الى قرى كبيرة وأخرى متوسطة وثالثة صغيرة (وهذا هو الاسلوب الطبقى فى المعاينة) ثم اختيرت من كل طبقة ٢٠٪ بحد أدنى قريتين من مجموعة القرى ...ة وقريتين الى ثلاث من القرى المتوسطة والكبيرة لكل محافظة ويتوقف ذلك درجة تجانس الطبقتين الاخيرتين ثم اختير من كل قرية (اختيرت فى المرحلة الاولى) العدد المقررمن الاسر بالطريقة المنتظمة وبذلك بلغ عدد القرى التى أختيرت فى المرحلة الاولى (العنقودية) ١١٤ قرية من جملة عدد قرى الجمهورية وعددها ٤٠٢٩ فى ذلك الحين .

ورغم ما يتميز به أسلوب المعاينة العنقودية بانخفاض التكاليف إلا أنه

يؤخذ عليه أن الوحدات المتجاورة فى وحدة المعاينة الاولية (المساكن فى الاحياء أو التجمعات السكنية مثلا) عادة ما تكون متشابهة فى الخصائص مما يعنى انها سوف لا تقدم معلومات اضافية تفيد الدراسة . كما أن اختيار عدد محدود من الوحدات – اذا لم يحسن اختيارها لتمثيل المجتمع – فقد يؤدى ذلك الى استنتاجات متحيزة .

وأخيرا وأيا كان أسلوب أو طريقة المعاينة الذى يتبع فلابد وأن يحقق الاسلوب الذى يختار ما يأتى :-

١. الكفاية من حيث النوع وذلك باختيار طريقة المعاينة التى تؤدى الى اختيار عينة ممثلة للمجتمع الاصلى أحسن تمثيل .

٢. الكفاية من حيث الحجم بحيث لا يكون حجم العينة أكبر من اللازم ويؤدى ذلك الى انفاق غير اقتصادى ولا أن يكون عدد الوحدات أقل من اللازم فلا يحقق الغرض وتكون التقديرات التى نحصل عليها من مثل هذه العينة أقل دقة عما ننشده .

وعادة ما يتحدد عدد الوحدات التى يجب أن تضمها العينة عدة عتبارات منها :-

١- مقدار الإعتمادات المخصصة للبحث . فكلما قلت الاعتمادات كلما استوجب ذلك تخفيض حجم العينة .

٢-حجم المجتمع الاصلى ودرجة تجانسه من حيث الظاهرة ومجموعة الظواهر موضوع الدراسة وكلما انخفضت درجة التجانس كلما أصبح من الضرورى زيادة حجم العينة .

الباب الثانى

التوزيعات التكرارية وتمثيلها بيانيا

Frequency Distributions & Graphical Representation

بعد اجراء عملية جمع البيانات سواء عن طريق العينة أو عن طريق الحصر الشامل ، يلاحظ أن أول مشكلة تواجه الباحث هى كيف يتسنى له عرض هذه البيانات بطريقة عملية تمكنه من التعرف على خصائص الظاهرة التى تعبر عنها البيانات [illegible] طيع دراس [illegible] ا ومقارنتها بالظواهر المشابهة لها . ومن البديهى ان أول طريقة من [illegible] ق ال [illegible] ض هو [illegible] ع هذه البيانات فى صورة جداول [illegible] ى معينة بحيث تفى بالغرض من [illegible] عرض البيانات بصورة ملخصة دون الأخلال بها فلا يكون التلخي [illegible] د بحيث تختفى المعالم الأصلية للبيانات ولا مستفيضا بحيث تعرض البيانات كما هى . وه [illegible] بعض [illegible] الملاحظات والشروط العامة التى يجب مراعاتها عند وضع البيانات الخام [illegible] جداول أولية تهينه لعرضها جدوليا بطرق احصائية نوجزها فيما يلى :

١. يجب أن يحتوى الجدول على كل البيانات التوضيحية الخاصة بالظاهرة التى يتم عرضها كما يجب أن يكون عنوان الجدول قصيرا دون غموضا وموضح فى الجدول متى وأين ولماذا جمعت هذه البيانات باختصار شديد .

٢. إذا كان رؤوس الجداول تعتمد على استخدام بعض الرموز المعبره عن مكونات الجدول ، فى مثل هذه الحالات يجب تذييل الجدول بتوضيح لهذه الرموز .

٣. طريقة ترتيب الاعمدة فى الجدول يجب أن يراعى مثلا التسلسل التاريخى إن وجد أو أهمية العمود فتكون الاعمدة الأولى هى المهمة ثم التى تليها وهى الأقل فى الأهمية إلى أن نأتى إلى آخر الجدول فيكون العمود الأخير هو عمود المجموع .

٤. فى الحالات التى يحتوى فيها الجدول على عدد كبير من الصفوف فيفضل ترك مسافة بين كل خمسة صفوف ن هذا يسهل عملية المقارنة والاستقراء والعرض .

٥. فى الأبحاث العلمية يراعى أن يكتب مصدر البيانات أسفل الجدول ، كما يراعى أيضا ذكر وحدات القياس ودرجة دقتها وهل هى مقربة أو كلها صحيحة .

أما وأن العرض البيانى والقياس الحسابى يعتمدان على طبيعة البيانات ونوع المتغير فإن ذلك يستوجب منا تناول البيانات أو المشاهدات ونوع القياسات التى تستخدم فى رصدها وانواع المتغيرات بالعرض المحدود دون الدخول فى تفصيلات أو علاقات رياضية قد تضفى على هذا الامر غموضا لا مبرر له .

(٢-١) أساليب القياس وأنواع البيانات :-

يعنى القياس رصد أو إعطاء أرقام أو اعداد تعبر عن المشاهدات بطريقة تسمح باستجلاء ما تعرضة هذه المشاهدات عن الظاهرة موضع الدراسة بإعداد جدول أو خريطة بيانية ثم إجراء عمليات حسابية تنتهى بحساب مقياس احصائى أو أكثر ، وهذا المقياس ملخص صفة أو اتجاها لهذه الظاهرة كتعيين قيمة تتركز حولها المشاهدات المسجلة عن تلك الظاهرة مثلا - واخيرا تنتهى الى تعميم او اتخاذ قرار فى شأنها .

وهناك أربعة أنواع أو مستويات رئيسية للقياس هى :

(٢-١-١) القياس النوعى أو التصنيفى

Nominal or Classificotory Scale

والذى يعنى رصد أرقام لتصنيف المتغير (أشياء - أشخاص - صفات ... الخ) فى مجموعات تشترك كلها فى صفة واحدة معينة . على سبيل المثال التصنيف النوعى لمجموعة من المترددين على مركز تجارى بحسب النوع (ذكور - إناث) أو حسب الحالة الأجتماعية (متزوج - أرمل - مطلق - لم يسبق له الزواج) وفى هذه الحالة فان التصنيف وكذلك القياس لابد وأن يتيح لكل مشاهده فرصة واحدة فقط للرصد دون الالتزام بأى ترتيب للفئات.

(٢-١-٢) القياس الترتيبى *Ordinal Scale*

يعنى هذا القياس بالإضافة الى المقياس السابق فان وحدة القياس لابد وأن تسمح بترتيب المشاهدات حسب مراكز متتابعة فى الأهميـة أو الدرجـة . فالتقديرات مثلا (ممتاز - جيد جدا - جيد - مقبول - ضعيف - ضعيف جدا) تمثل مقياسا ترتيبيا ولا يشترط أن تكون فئات المتغير التى تنشأ عن هذا الاسلوب متساوية الطول .

(٢-١-٣) القياس بفترة *Interval Scale*

والمقياس فى هذه الحالة برصد قيمـا أو أعداد حقيقية للمشاهدات بحيث تكون هناك مسافات متساوية بين وحدات القياس وهذه المسافات المتساوية تعنى وجود مسافات متساوية بين قيم المتغير المقيس هذا بالاضافة الى أن نقطة الاصل فى هذا المقياس تحكمية . فمثلا فى درجات الحرارة مقاسة بالفهرنهيت فإن درجة الحراره ٥٠ مثلا أعلى من درجة الحرارة ٤٠ وأن الفترة بين درجتى الحرارة ٥٠ ، ٦٠ هى نفس الفترة بين درجتى الحراره ٦٠ ، ٧٠ درجة ، كما أن نقطة الاصل ٣٢ هى نقطة أصل تحكمية .

(٢-١-٤) القياس بالنسبة *Ratio Scale*

طبقا لهذا المقياس فإن القيم التى تعطى للمشاهدات لها نفس خصائص [illegible] القياس بفترة بالاضافة الى أن نقطة الاصل هنا هى الصفر [illegible]

الخاصية المقيسة ، فمثلا العمر يبدأ بالصفر عند المولد والعمر ٦٠ عاما يزيد بنفس العدد من السنوات عن العمر ٥٠ عاما ولأن الصفر هى نقطة الأصل فى هذه الحالة فإن العمر ٨٠ يعنى ضعف العمر عند السن ٤٠ على العكس من القياس بفترة (نقطة الاصل تحكمية) .

وكما سيتبين لنا فى هذا الباب وأبواب أخرى من هذا الكتاب فإن أسلوب العرض البيانى وكذلك المقاييس الاحصائية بل وأنواع الاختبارات الأحصائية تختلف باختلاف المتغير ووحدات القياس التى رصدت على اساسها المشاهدات المسجلة والخاصة بهذا المتغير .

وهناك نقطة أخيره نرد أن نذكرها ونحن فى هذا المجال وهى أن القياس بفترة أو بنسبة وبالتالى المشاهدات المسجلة على اساس هذين المقياسين قد تعبر عن أحد متغيرين : إما أن يكون المتغير مستمرا (*Continuous*) وفى هذه الحالة فإن المشاهدات التى تسجل عن هذا المتغير يمكن ان تأخذ (قيم صحيحة أو كسرية) أى قيمة داخل المدى (وعددها لانهائى) الذى يحدد جميع القيم الممكنة للمتغير ومن أمثلة هذا المتغير المشاهدات (البيانات) المسجلة عن درجة الحرارة وطول القامة والوقت . وإما أن يكون المتغير متقطع (*discrete*) أو غير مستمر وفى حالتنا هذه فان المتغير يأخذ قيما محدودة (أى قيم صحيحة فقط) داخل المدى والأمثلة على ذلك عديدة منها المتغير الذى يعبر عن عدد أفراد الاسرة وعدد الحجرات بالوحدة السكنية وعدد الاخطاء المطبعية فى كتاب.

وقبل أن ننتقل الى الفقرة التالية فإنه يتعين علينا ملاحظة أن البيانات (المشاهدات الخاصة بالمتغيرين المستمر والمتقطع تسمى بيانات كمية ، اما البيانات المسجلة على اساس المقياسين النوعى والترتيبى فيطلق عليها بيانات وصفية .

ومن وجهة أخرى يمكن تقسيم البيانات الى بيانات بسيطة وبيانات مزدوجة ، فالبيانات البسيطة هى بيانات خاصة بظاهرة واحدة سواء كانت هذه الظاهرة كمية أو وصفية ، أما البيانات المزدوجة فهى خاصة بظاهرتين بينهما علاقة ، فمثلا قد يكون لدينا مجموعة [illegible] [illegible]اج وأعمار زوجاتهم (علاقة ظاهرة كمية بأخرى كمية) أو تقديرات مجموعة من [illegible] (علاقة ظاهرة وصفية [illegible] يكون احدى [illegible] كمية والأخرى وصفية .

وتتم عملية تبويب البيانات وعرضها باحدى الطريقتين الآتيتين :

- تبويب البيانات وعرضها فى صورة جداول إحصائية
- تبويب البيانات وعرضها فى صورة خرائط أو رسوم بيانية

(٣-٢) العرض الجدولى للبيانات :-

يقصد بالعرض الجدولى للبيانات الاحصائية وضعها فى توزيعات (أو جداول) تمكن من الحصول على المعلومات بصورة سهلة وواضحة ومختصرة تسهل من دراسة الظاهرة موضوع الدراسة .

ومن الطبيعى ان تختلف الجداول التكرارية بحسب البيانات الى يراد تبويبها وما اذا كانت بيانات وصفية أو كمية ، وهذا ما سنلاحظة عند استعراض الانواع المختلفة للجداول الاحصائية .

(٢-٣-١) أنواع الجداول الاحصائية :

حيث تتمثل الجداول الاحصائية فى جداول احصانية بسيطة وجداول مركبة وأخرى تكرارية .

• الجداول الاحصائية البسيطة *Simple Table*

فى هذا النوع من الجداول يتم توزيع المفردات بحسب الحالات المختلفة لظاهر ، ما تتكون هذه الجداول من عمودين أحدهما يمثل الحالات الممكنة للظاهرة والثانى يمثل عدد مفردات كل حالة، مثال ذلك بيان الكميات المنتجة من سلعة ما خلال الفترة ١٩٩٠ - ١٩٩٤ .

السنة	١٩٩٠	١٩٩١	١٩٩٢	١٩٩٣	١٩٩٤
الكمية المنتجة بالطن	٤	٣	٦	٧	٨

• ثانيا الجداول المركبة Double Table

وهى تلك التى تشمل توزيع المفردات بحسب الحالات المختلفة لاكثر من متغير واحد ، ومثال ذلك بيان الكميات المنتجة والمبيعات من سلعة ما خلال الفترة ١٩٩٠ - ١٩٩٤.

السنة	١٩٩٠	١٩٩١	١٩٩٢	١٩٩٣	١٩٩٤
الكمية المنتجة بالطن	٤	٣	٦	٧	٨
المبيعات بالطن	٥	٢	٦	٦	٩

• الجداول التكرارية Frequency Table

اذا كانت البيانات الاحصائية عبارة عن عدد كبير من المفردات التى تعبر عن مقاييس كمية لظاهرة معينة فإن هذه المفردات تقسم الى مجموعات جزئية حسب خاصية معينة ثم يتم ترتيبها تصاعديا أو تنازليا بحيث يطلق على كل منها "فئة" وكذلك فإن عدد المفردات فى كل فئة يطلق عليه تكرار فى هذه الحالة يسمى الجدول الذى يضم كل من الفئات والتكرارات المناظر اسم التوزيع التكرارى وهذا التوزيع التكرارى هو ما سيتم الاعتماد عليه فى دراسة اى ظاهرة فى الابواب التالية .

* أنواع الجداول التكرارية :-

من المعروف ان الجداول التكرارية يتم تقسيمها الى نوعين هما الجداول التكرارية البسيطة والمزدوجة ، وسوف نعرض بإيجاز مناسب لهذه الانواع من الجداول التكرارية فى حالتى البيانات الكمية والوصفية .

١ – الجداول التكرارية البسيطة (*One- way Frequency Table*)

للبيانات الكمية كما سبق وأن ذكرنا فإنه يوجد نوعان من المتغيرات الكمية هى المتغيرات المتقطعة (المنفصلة) وهى التى لا تأخذ إلا قيما صحيحة فقط مثل متغير عدد الغرف والوحدات السكنية وعدد الابناء الاحياء وكذلك أى متغير تعتمد قيمته على العد الطبيعى للاشياء ، والمتغيرات المستمره (المتصلة) وهى التى تأخذ جميع القيم الممكنة صحيحة أو كسرية موجبة أم سالبة مثل متغير العمر والطول والوزن وغيرها من المتغيرات التى تعتمد على القياس من أى نوع ، فطول الانسان يمكن ان يكون ١٧٦,٥سم أو أى عدد من السنتيمترات وأجزاء من السنتيمتر مهما كان الجزء صغيرا .

وأولى خطوات التبويب لأى متغير هى حصر القيم المختلفة للمتغير وهى التى تمثل الاجابة على السؤال المناظر للمتغير فى استماره البحث ، يلى ذلك تفريغ البيانات فى الجدول وذلك بوضع علامة امام القيم المناظرة ثم نقوم بعد هذه العلامات لنحصل على التكرار المطلوب كما يتضح من المثال التالى .

مثال (٢) :-

الاتى بيانات ٢٠ أسرة وعدد غرف السكن لها ، المطلوب انشاء الجدول التكرارى المناسب :

٤، ١، ٣، ٢، ٢، ١، ٢، ٤، ٥، ٣، ٢، ٢، ١، ٥، ٣، ٢، ٢، ٣، ٣

الاجابة

جدول تفريغ بيانات الاسر وعدد غرف السكن

عدد الغرف	العلامات	عدد الاسر
١	////	
٢		٧
٣	/////	-
٤	//	
٥	//	٢
المجموع	—	٢٠

أما فى حالــة المتغيرات المتصلة وحيث انهــا تـأخذ أى نـوع مـن القيـم فأننا نقـوم بتجميع مجموعات القيم المتقاربـة فيما يســمى بفئات قيـــم المتغير (*Class Intervals*) ثم نقوم بتفريغ البيانات بوضع علامات القيم أمـام الفئـات

المتناظرة ثم نحول هذه العلامات الى تكرارات حسب الطريقة المشروحة سابقا فنحصل على التوزيع التكرارى المطلوب .

وهناك عده اعتبارات يجب مراعاتها عند تكوين فنات الجدول التكرارى نوردها فيما يلى :

١.يجب الا يكون عدد الفنات صغيرا فنفقد بذلك معالم التوزيع الحقيقى أو كبيرا جدا فينتفى الهدف من تكوين الجدول وهو الاختصار ، وقد جرى العرف على الايزيد عدد الفنات عن ٢٠ فنة ولا يقل عن ٦ ويتوزع بين ذلك حسب حجم البيانات وطريقة توزيع القيم . وعموما يمكن تحديد عدد الفنات عن طريق تطبيق العلاقة التالية :

عدد الف[illegible] ٣,٣ لوغاريتم عدد القيم (او مجموع التكرارات)

تحديد طول الفنة أ[illegible] [illegible] التى تنحصر فيه قيم المتغير بالنسبة لهذه الف[illegible] . ولا يلزم ان تكون أطوال الفنات متساوية وإن كان تحقيق هذا الشرط يؤدى الى سهولة العمليات الحسابية التى تجرى على الجدول فيما بعد .

وعموما ولكى نحسب طول الفنة فلابد أولا من حساب المدى الذى تنحصر فيه جميع قيم المتغير (المدى = اكبر قيمة للمتغير – أصغر قيمة للمتغير) ، يلى ذلك تحديد طول الفنة وذلك بقسمة المدى على عدد الفنات الذى سبق تحديده ، ويمكن عمل العكس أى تحديد طول الفنة

اولا ثم قسمة المدى عليها فنحصل على عدد الفئات المناسب . ومن المعلوم أنه يمكن استخدام التقريب اذا كانت نتيجة القسمة عددا كسريا ويمكن أن نعدل حدود الفئات بحيث تمثل هذه الحدود قيما يسهل تذكرها والتعامل معها .

٣.نقوم بكتابة الفئات فى العمود الاول فى جدول التفريغ بطريقة تمنع التداخل والازدواج بين الفئات المختلفة أو تؤدى إلى سقوط بعض القيم وعدم وضعها فى فئات مناسبة وأيضا بطريقة توضح الحد الادنى والحد الأعلى لكل فئة . فإذا كان لدينا فئات متساوية الطول وهو خمسة قيم وإذا كانت بداية الفئة الاولى القيمة ١٥ فإن هذه الفئات تكتب كما يلى :

الفئة الاولى :	من ١٥ الى	أقل من ٢٠
الفئة الثانية :	من ٢٠ الى	اقل من ٢٥
الفئة الثالثة :	من ٢٥ الى	أقل من ٣٠
،	،	،
،	،	،
،	،	،

وهكذا . . . . .

ويمكن اختصار كتابة الفئات على النحو التالى لتوفير المساحة :

١٥ – .

٢٠ –

٢٥ –

حيث نكتفى بكتابة الحد الادنى للفئة ونضع بجانبة شرطة أفقية معناها الى اقل من الحد الأعلى .

ثم نستكمل باقى اجراءات تكوين الجدول التكرارى حسب القواعد السابقة.

مثال (٣) :–

انشأ الجدول التكرارى الذى يلخص درجات ١٠٠ طالب فى مادة الاحصاء والدرجات هى :–

٧٥	١١	٤٣	٨٥	٩٠	٤٥	٤٦	٧٠	٥٩	١٥
٩٥	٥٣	٥٤	٧٢	٨٥	١٨	٤٩	١٢	١٧	٩٦
٣٢	٤٥	٤٠	٨١	٥١	٢٨	٢٢	٦٢	٥٦	٧٥
٥٣	٨٤	٤٠	٥٨	٥١	٧٨	٤٧	٤٨	٥٠	٥٣
٤٣	٢٧	٦٧	٦٦	٣٩	٢٩	٨٠	٥٥	٦٣	٦٤
٧٦	٨٤	٧٦	٢٧	٦١	١٩	٦٣	٨٠	٦٩	٧٥
٥٣	٣٨	٣٧	٤٥	١٢	٥٧	٥٧	٤٧	٧٧	٤٤
١٦	٤١	١٣	٢٩	٨٢	٧٢	١٩	٨٦	٧٢	٣٥
٢٢	٩١	٣٥	٥١	٢٦	١٧	٧٧	٥٨	٧٢	٧٤
١٧	٥٤	٦٢	٤٩	[illegible]	٧١	٨٠	١٤	٣٧	٦٦

الحل

أولا : تحديد عدد الفئات :

عدد الفئات = ١ + ٣,٣ لو ١٠٠ = ٨ فئات (تقريبا)

ثانيا : تحديد مدى البيانات:

المدى = اكبر مفردة – اصغر مفرده

= ٩٥ – ١١ = ٨٤

ثالثا : [illegible] الفئة :

طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$

= $\frac{١٤}{٨}$ ١١

رابعا : ننشأ الجدول التكرارى كالاتى :

الجدول التكرارى لدرجات الطلاب

الفئات	العلامات	التكرارت
١١ -	~~////~~ ////	٩
٢٢ -	~~////~~ ~~////~~ /	١١
٣٣ -	~~////~~ ~~////~~ ///	١٣
٤٤ -	~~////~~ ~~////~~ ~~////~~ ////	١٩
٥٥ -	~~////~~ ~~////~~ ////	١٤
٦٦ -	~~////~~ ~~////~~ ~~////~~ /	١٦
٧٧ -	~~////~~ ~~////~~ ///	١٣
٨٨ - ٩٩	~~////~~	٥
مجموع	—	١٠٠

*** ملاحظات على الجدول التكرارى :-**

١.الجدول التكرارى السابق جدول منتظم (أطول فئات الجدول جميعها متساوية) أما إذا اختلفت أطوال الفئات ، فيقال ان الجدول غير منتظم.

٢.الجدول التكرارى السابق مغلق وذلك لتحديد بداية الفئة الاولى ونهاية الفئة الاخيرة صراحة بالجدول ، أما اذا لم يتحدد بالجدول كل من بداية الفئة الاولى ونهاية الفئة الأخيرة أو كلاهما فيقال ان الجدول التكرارى

مفتوح من أعلى (بداية الفئة الاولى غير محدده) أو مفتوح من اسفل (نهاية الفئة الاخير غير محدده) أو مفتوح من الطرفين (بداية الفئة الاولى ونهاية الأخيرة غير محددتين)

٣. يحتوى الجدول التكرارى على كل البيانات سواء أصغر القيم المفرغة أو أكبرها .

٤. أخيرا يلاحظ دائما فى الجداول التكرارية أن :

مجموع التكرارات = عدد بيانات هرة .

وعادة ما يستبعد العمود الثانى المخصص للعلامات لانه يعتبر فقط وسيلة لمعرفة عدد التكرارت داخل كل فئة ، وبذلك نحصل على وزيع التكر لتوزيع درجات الطلاب :

التوزيع التكرارى لدرجات الطلاب

فئات الدرجة	التكرار (عدد الطلاب المناظر)
١١ -	٩
٢٢ -	١١
٣٣ -	١٣
٤٤ -	١٩
٥٥ -	١٤
٦٦ -	١٦
٧٧ -	١٣
٨٨ - ٩٩	٥
المجموع	١٠٠

وعلى العموم فإن توزيع المفردات الاصلية على الفئات المختلفة فى أى جدول تكرارى يؤدى الى اختفاء القيم الاصلية ويضيع معالمها ، اى أنه لا يمكننا معرفة القيمة الاصلية لأى مفرده من الجدول ولا نعرف عنها إلا أنها واحد فى فئة معينة بين حدود معينة . ففى الجدول السابق لا نعرف شيئا عن الدرجات الأصلية فى الفئة الاولى سوى أن عددها ٩ ولا ندرى ما إذا كانت كلها أو بعضها متساوية وما اذا كانت كلها قريبة من ١١ أو ٢٢ أو تقع كلها فى منتصف الفئة ، لذلك

٩ ؛

فنحن نفترض دائما فى الجدول التكرارى أن التكرارات فى كل فئة لها نفس قيمة مركز الفئة .

ويعرف مركز الفئة (*Class Midpoint*) بأنه القيمة التى تقع فى منتصف الفئة ويحسب على النحو التالى :

مركز الفئة = (الحد الاعلى للفئة + الحد الادنى للفئة) ÷ ٢

ولاسباب معي[illegible] يصعب معها تحديد الحد الادنى للفئة الاولى (توزيع تكرارى مفتوح [illegible] أعلى) وأيضا الحد الاعلى للفئة الأخيرة (توزيع تكرارى مفتوح من أسفل) أو [illegible] معا ، فى مثل هذه الحالات يأخذ عمود الفئات فى هذه الحالة الاشكال التالية :

أقل من ١٠	أو	صفر – ١٠	و	أقل من ١٠
١٠ – ٢٠		١٠ – ٢٠		١٠ – ٢٠
٢٠ – ٣٠		٢٠ – ٣٠		٢٠ – ٣٠
٣٠ – ٤٠		٣٠ فأكثر		٣٠ فأكثر

واضح أيضا أنه يصعب علينا تحديد مركز الفئة المفتوحة مما سيترتب عليه بعض المصاعب فى تمثيل هذه الجداول بيانيا وفى حساب بعض المقاييس كما سنرى فيما بعد .

وللتسهيل تم الاتفاق على اهمال الارقام الموجودة على يسار العلامة (–) فى كل فئة من فئات التوزيع التكرارى ما عدا الفئة الاخيرة وبذلك يكون مفهوم ضمنا أن الفئة الاولى تبدأ من الحد الادنى للفئة الاولى وتنتهى قبل بداية الفئة الثانية وهكذا وبذلك يمكن كتابة الفئات كالاتى أيضا :

صفر–　　　١٠ –　　　٢٠ –　　　٣٠ – ٤٠

<u>* تبويب المتغيرات الوصفية :–</u>

اولى خطوات التبويب لأى متغير وصفى هى حصر الحالات أو الصفات التى يأخذها المتغير والتى تمثل الاجابات الممكنة على السؤال المناظر للمتغير فى استمارة البحث ، ثم يتم اتباع نفس خطوات تكوين الجدول التكرارى السابق ذكرها فى حالة البيانات الكمية ، وبالمثل يمكن القول بأنه لو استبعدنا العمود الخاص بالعلامات فسوف نحصل على التوزيع التكرارى .

مثال (٤) :-

الاتى بيانات ٢٥ شخصا بالنسبة للحالة الزواجية والمطلوب عرضها فى صورة جدول تكرارى :

أعزب ،	أعزب ،	مطلق ،	مطلق ،	أعزب ،
أرمل ،	متزوج ،	اعزب ،	متزوج ،	متزوج
متزوج ،	ارمل ،	مطلق ،	متزوج ،	ارمل
أعزب ،	أرمل ،	مطلق ،	متزوج ،	أعزب
متزوج ،	أعزب ،	متزوج ،	اعزب ،	مطلق

الحل

بحصر حالات المتغير نجد أن لدينا أربع حالات هى أعزب ، متزوج ، مطلق ، أرمل . فنقوم بتفريغ بيانات الـ ٢٥ شخصا حسب هذه الحالات فى جدول التفريغ - حيث يذكر فى العمود الاول الحالات الأربع على الترتيب ويوضح اسم المتغير على رأس العمود ثم نقوم بوضع العلامات فى العمود الأوسط ، ثم نقوم بعد العلامات امام كل صفة ويرمز ذلك العدد الى التكرار وبذلك يمكن الحصول على جدول التوزيع التكرارى على الصور التالية :

شكل (١) جدول تفريغ توزيع الأشخاص حسب الحالة الزواجية

الحالة الزواجية	العلامات	التكرار
أعزب	~~////~~ ///	٨
متزوج	~~////~~ ///	٨
مطلق	~~////~~	٥
أرمل	////	٤
المجموع	----	٢٥

شكل (٢) جدول التوزيع التكرارى

الحالة الزواجية	التكرار
أعزب	٨
متزوج	٨
مطلق	٥
أرمل	٤
المجموع	٢٥

٢ – الجداول التكرارية المزدوجة *Two - Way Frequency Table*

تستخدم التوزيعات التكرارية البسيطة لتمثيل ظاهرة واحده ، اما اذا كنا نود دراسة ظاهرتين أو ظاهرة واحدة تضم متغيرين فى الوقت نفسه مثل طول الاب وطول الابن أو كمية المبيعات من احدى السلع ومصاريف الاعلان عن السلعة ... الخ ، فانه يمكن اتباع نفس الخطوات السابقة تقريبا لاعداد جدول تكرارى مزدوج لتمثيل الظاهرتين ، وفيه تخصيص الاعمدة لتوزيع احدى الظاهرتين ، بينما تخصص الصفوف لتوزيع الظاهرة الثانية عند بدايات الصفوف وتتبع نفس الاجراءات السابقة لتحديد عدد واطوال الفئات لكل ظاهرة ، وعند تفريغ المفردات نأخذ أزواج القيم للظاهرتين بدلا من قيمة واحدة فى حالة الجدول

التكرارى البسيط ونضع لكل قيمتين متناظرتين علامة (شرطة مائلة) فى الخلية التى تقابل فئتيهما .

مثال (٥) : -

فيما يلى البيانات الخاصة بالطول بالسنتيمتر (س) والوزن بالكيلو جرام (ص) لعدد ٢٠ شخص .

المسلسل	الطول (س)	الوزن (ص)	المسلسل	الطول (س)	الوزن (ص)
١	١٨٩	٨٤	١١	١٨٠	٧٢
٢	١٦٠	٥٨	١٢	١٧٠	٦٥
٣	١٨٧	٧٨	١٣	١٧٧	٧٧
٤	١٦٢	٥٩	١٤	١٧٠	٦٨
٥	١٨٥	٧٦	١٥	١٧٧	٧٠
٦	١٦٥	٦١	١٦	١٧٠	٧٢
٧	١٨٣	٧٥	١٦	١٧٥	٧٦
٨	١٦٥	٦٧	١٨	١٧٣	٦٧
٩	١٨٠	٧٩	١٩	١٧٥	٧٢
١٠	١٦٨	٦٣	٢٠	١٧٥	٦٩

والمطلوب : انشاء الجدول التكرارى المزدوج لطول ووزن هؤلاء الأشخاص

الحل

لعمل الجدول التكرارى المزدوج لتمثيل ظاهرتى الطول (س) والوزن (ص) نحدد أولا مدى التغير لكل من الظاهرتين . ثانيا نقسم كل مدى على عدد الفئات المناسبة (كما سبق شرحه فى عمل الجداول التكرارية البسيطة) . ثالثا نقوم بتفريغ البيانات فى جدول التفريغ المزدوج وذلك بأن نضع علامة لكل قيمتين متناظرتين فى الخلية التى تقابل فئتيهما .

وبذلك يكون الحل كما يلى :

بالنسبة لظاهرة الطول (س) :-

١ - المدى = ١٨٩ - ١٦٠ = ٢٩

٢ - عدد الفئات = ١ + ٣,٣ لو مجموع التكرارات
= ١ + ٣,٣ لو ٢٠ = ٦ فئات تقريبا .

٣ - طول الفئة المناسب = $\frac{\text{المدى}}{\text{عدد الفئات}}$ = $\frac{٢٩}{٦}$ = ٥ سنتيمتر تقريبا

فتكون فئات ظاهرة الطول (س) هى : ١٦٠- ، ١٦٥- ، ١٧٠- ، ١٧٥- ، ١٨٠- ، ١٨٥-١٩٠

بالنسبة لظاهرة الوزن (ص) :-

١ - المدى = ٨٤ - ٨٥ = ٢٦ كيلو جرام

٢ - عدد الفئات = ١ + ٣,٣ لو ٢٠ = ٦ فئات تقريبا

٣ - طول الفئة المناسب = $\frac{٢٩}{٦}$ = ٥ كيلو جرام تقريبا

فتكون فئـات ظـاهرة الـوزن (ص) هـى:٥٥- ، ٦٠- ، ٦٥- ، ٧٠- ، ٧٥- ، ٨٠-٨٥

وعند عمل جدول التفريغ المزدوج تتبع الطريقة الآتية :

الشخص الأول مثلا طوله ١٨٩سم ، ووزنه ٨٤ كجم فنضع علامـة تمثـل هذا الزوج من القيم فى الخانة المقابلة للفئتين (١٨٥-) للطول ، (٨٠-) للوزن .

وهكذا بالنسبة لباقى القيم وبذلك يكون جدول التفريغ المزدوج كما يلى :

شكل (١) تفريغ بيانات الاطوال والاوزان

(ص) / (س)	٥٥-	٦٠-	٦٥-	٧٠-	٧٥-	٨٠-٨٥	المجموع
١٦٠-	//						//
١٦٥-		//	/				///
١٧٠-			///	/			////
١٧٥-			/	//	//		~~////~~ /
١٨٠-				/	//		///
١٨٥-١٩٠					//	/	///
المجموع	//	//	~~////~~	////	~~////~~ /	/	~~////~~ ~~////~~ ~~////~~ ~~////~~

وبحصر عدد العلامـات فـى جـدول التفريـغ المـزدوج نحصـل علـى الجـدول التكرارى المزدوج المطلوب كما يلى :

شكل (٢) الجدول التكراري المزدوج لطول ووزن ٢٠ شخص

(ص) / (س)	٥٥-	٦٠-	٦٥-	٧٠-	٧٥-	٨٠-٨٥	المجموع
١٦٠-	٢						٢
١٦٥-		٢	١				٣
١٧٠-			٣	١			٤
١٧٥-			١	٢	٢		٥
١٨٠-				١	٢		٣
١٨٥-١٩٠					٢	١	٣
المجموع	٢	٢	٥	٤	٦	١	٢٠

ونود أن ننوه في هذا المجال انه يمككنا الحصول على التوزيع التكراري الهامشي لكل الظاهره (س) ، الظاهره (ص) وذلك بفصل فئات كل متغير والتكرارات المناظره له في جدول مستقل .

فمثلا نجد أن التوزيع الهامشي لظاهرة الطول (س) دون أخذ ظاهره الوزن (ص) في الاعتبار يأخذ الصوره الاتيه :

التوزيع الهامشى لتغير الطول

فئات الطول (س)	التكرار
-١٦٠	٢
-١٦٥	٣
-١٧٠	٤
-١٧٥	٥
-١٨٠	٣
١٩٠-١٨٥	٣
المجموع	٢٠

كذلك فان التوزيع الهامشى لظاهرة الوزن (ص) دون اخذ ظاهره الطول (س) فى الأعتبار يأخذ الصوره التالية :

التوزيع الهامشى لتغير الوزن

فئات الطول (ص)	التكرار
-٥٥	٢
-٦٠	٢
-٦٥	٥
-٧٠	٤
-٧٥	٦
٨٥-٨٠	١
المجموع	٢٠

وكما هو معلوم فى حالة الجداول التكرارية البسيطة ، فإن الجداول التكرارية المزدوجة لا تقتصر على البيانات الكمية فقط بل يمكن انشأ جداول تكرارية مزدوجة للبيانات الوصفية أيضا ، ومن المعلوم أيضا وجود نوعان من الجداول التكرارية المزدوجة فى حالة الظواهر الوصفية هما :

- **جدول الاقتران :**

حيث يمثل هذا الجدول ظاهرتين وصفيتين كل منهما قسمين فقط. فإذا فرضنا مثلا وجود ٤٠ طالب مقسمين حسب الجنس (ذكور – اناث)، وحسب التخصص (علمى – أدبى) ، فى هذه الحالة يأخذ جدول الاقتران الصورة التالية :

التخصص / الجنس	علمى	أدبى	المجموع
ذكور	١٥	١٠	٢٥
اناث	١٠	٥	١٥
المجموع	٢٥	١٥	٤٠

• جدول التوافق :

يمثل هذا النوع من الجداول التكرارية المزدوجة العلاقة بين ظاهرتين يمكن تقسيم اى منهما أو كليهما الى أكثر من قسمين . فاذا فرضنا ان ١٠٠ شركة قسمت من حيث مستوى الربحية الى عالية الربحية ، منخفضة الربحية ، منعدمة الربحية ، منخفضة الخسائر ، عاليةالخسائر ، ومن حيث الحجم الى شركات كبيرة وشركات صغيرة وبالتالى فإن جدول التوافق يأخذ الشكل التالى :

جدول التوافق

مستوى الربحية / الحجم	عالية الربحية	منخفضة الربحية	منعدمة الربحية	منخفضة الخسائر	عالية الخسائر	المجموع
كبيرة	٢٠	١٨	١٢	٨	٤	٦٢
صغيرة	١٢	١٠	٨	٢	٦	٣٨
المجموع	٣٢	٢٨	٢٠	١٠	١٠	١٠٠

٣ – الجداول التكرارية المتجمعة :

رأينا فيما سبق أن التوزيع التكرارى هو ذلك الجدول الذى يحتوى على عمودان الاول يمثل فئات الظاهرة والثانى يمثل التكرار المناظر لكل فئة ، ولكننا فى بعض الاحيان قد نرغب فى معرفة عدد مفردات الظاهره التى تقل أو تزيد عن قيمة معينة ضمن قيمة الظاهرة . كذلك تحديد نسبة مفردات الظاهرة التى تقل أو

تزيد عن حد معين . الا أن هذه المشكلة يمكن حلها عن طريق تكوين ما يسمى بالتوزيع التكرارى المتجمع . ومن المعروف انه يوجد نوعان من التوزيعات التكرارية المتجمعة هما :

- التوزيع التكرارى المتجمع الصاعد : وهو عباره عن جدول مكون من عمودين اضافيين يمكن اشتقاقهما من عمودى التوزيع التكرارى الاصلى (عمود الفئات وعمود التكرار) أما العمود المشتق الاول فهو يمثل الحدود العليا لفئات الظاهرة مسبقة بكلمة أقل من والثانى يشمل التكرار المتجمع الصاعد ويبدأ بتكرار الفئة الاولى ثم يتم جمع تكرار كل فئة على مجموع التكرارات السابقة له بحيث تكون قيمة التكرار المتجمع الصاعد للفئة الأخيرة عبارة عن مجموع قيم تكرارات الظاهرة .

على سبيل المثال ، يلاحظ أن التوزيع التكرارى المتجمع الصاعد لجدول التوزيع الهامشى لمتغير الطول والذى تكوينه فى مثال (٥) يكون على الصورة الاتية :

جدول (١) جدول التوزيع التكرارى الاصلى — جدول (٢) جدول التوزيع التكرارى المتجمع الصاعد

فئات الطول (س)	التكرار	حدود عليا للفئات	تكرار متجمع صاعد
١٦٠ -	٢	أقل من ١٦٥	٢
١٦٥ -	٣	أقل من ١٧٠	٥
١٧٠ -	٤	أقل من ١٧٥	٩
١٧٥ -	٥	أقل من ١٨٠	١٤
١٨٠ -	٣	أقل من ١٨٥	١٧
١٨٥ - ١٩٠	٣	أقل من ١٩٠	٢٠
المجموع	٢٠		

* بعض الملاحظات على جدول التوزيع التكرارى المتجمع الصاعد :-

من أهم هذه الملاحظات انه فى حالة الجداول التكرارية المغلقة (بداية الفئة الأولى ونهاية الفئة الأخيرة) محدده صراحة فى الجدول التكرارى الأصلى فلا توجد صعوبات فى تكوين مفردات جدول التوزيع التكرارى المتجمع الصاعد . أما فى حالة الجداول التكرارية المفتوحة وخاصة من اسفل (نهاية الفئة الأخيرة غير محدده صراحة فى الجدول الأصلى) فى مثل هذه الحالات يتم الاستعاضة بعبارة أقل من (ص) للفئة الأخيرة وتكتب قيمة التكرار المتجمع الصاعد المناظر للفئة الأخيرة كما هو فى الحالة العادية .

يمكن الاعتماد على مثل هذه التوزيعات فى تحديد عدد المفردات (نسبة المفردات) التى تقل عن حد معين ، فمثلا يمكننا من خلال الجدول المتجمع السابق استنتاج انه يوجد ١٤ شخصا أطوالهم اقل من ١٨٠سم ، أو أنه توجد نسبة ٢٥٪ = ($\frac{٥}{٢٠} \times \frac{١٠٠}{١٠٠}$) من الاشخاص اطوالهم تقل عن ١٧٠سم .

• التوزيع التكرارى المتجمع الهابط : ويتكون هذا الجدول أيضا من عمودان الاول يشمل الحدود الدنيا لفئات الظاهرة متبوعة بكلمة فأكثر ، أما العمود الثانى فيمثل التكرار المتجمع الهابط والذى يبدأ بمجموع التكرارات أمام الفئة الاولى ثم يطرح من المجموع تكرار الفئة الأولى فى التوزيع الأصلى وذلك للحصول على التكرار المتجمع الهابط للفئة الثانية ثم يطرح من تكرار الفئة الثانية لنحصل على التكرار المتجمع الهابط المناظر للفئة الثالثة وهكذا حتى نصل إلى التكرار المتجمع الهابط المناظر للفئة الأخيرة ويكون مساويا لتكرارها الأصلى .

فالجدول التكرارى المتجمع الهابط للجدول التكرارى بالمثال السابق يأخذ الصوره :-

جدول (١) جدول (٢)

جدول التوزيع التكرارى الأصلى جدول التوزيع التكرارى المتجمع الهابط

فئات الطول	التكرار	حدود دنيا للفئات	تكرار متجمع هابط
١٦٠ -	٢	١٦٠ فأكثر	٢٠
١٦٥ -	٣	١٦٥ فأكثر	١٨
١٧٠ -	٤	١٧٠ فأكثر	١٥
١٧٥ -	٥	١٧٥ فأكثر	١١
١٨٠ -	٣	١٨٠ فأكثر	٦
١٨٥ - ١٩٠	٣	١٨٥ فأكثر	٣
المجموع	٢٠		

من هذا التوزيع أيضا يمكن إستنتاج أن هناك مثلا ١١ شخصا تزيد أطوالهم عن ١٧٥ سم أي أن ٧٥٪ ($\frac{١٥}{٢٠} \times \frac{١٠٠}{١٠٠}$) من أشخاص العينة أطوالهم تزيد عن ١٧٠سم وهكذا .

ومما يجب ملاحظته فى هذا المجال أن التوزيعات التكرارية بنوعيها الصاعدة والهابطة لا تتأثر مطلقا بانتظام أو عدم انتظام الجداول فيمكن إيجاد التوزيع التكرارى المتجمع الصاعد للجدول الغير منتظم بنفس الاسلوب ، كما أن التوزيعات التكرارية المتجمعة سواء الصاعدة منها أو الهابطة لا تتأثر بفتح الجدول أو إغلاقه .

وعادة ما تفيد هذه التوزيعات أيضا فى التمثيل البيانى للظواهر المختلفة . كما أنها تستخدم فى حساب بعض المقاييس الاحصائية مثلا الوسيط والربيعين الأدنى والأعلى ، كما سيتضح لنا فيما بعد .

٤ – الجداول التكرارية النسبية :–

لكى نتمكن من مقارنة توزيعين تكراريين ، يجب أن يكون المجموع الكلى لتكرارات التوزيعين متساويا حتى تكون المقارنة دقيقة ، أما اذا كان المجموعان الكليان للتكرارات غير متساويين فإن المقارنة بين التوزيعين مباشرة تعد مضللة ويجب أن تجرى فى هذه الحالة على التوزيع التكرارى النسبى ، إذ يتم تحويل التكرارات الى نسب عادية أو مئوية وذلك قبل إجراء عملية المقارنة حيث :

$$\text{التكرار النسبى} = \frac{\text{التكرار الأصلى}}{\text{مجموع التكرارات}}$$

والجدول الذى يحتوى على التكرارات النسبية يسمى التوزيع التكرارى النسبى فالتوزيع التكرارى النسبى للجدول التكرارى الذى تم تكوينه فى المثال السابق يمكننا الحصول عليه وذلك بقسمة كل تكرار فى الجدول على المجموع الكلى لتكرارات الجدول وهو :

جدول التوزيع التكرارى النسبى

فئات الطول	التكرار الأصلى	حدود عليا للفئات
١٦٠ -	٢	١٠٪
١٦٥ -	٣	١٥٪
١٧٠ -	٤	٢٠٪
١٧٥ -	٥	٢٥٪
١٨٠ -	٣	١٥٪
١٨٥ - ١٩٠	٣	١٥٪
المجموع	٢٠	١٠٠٪

ويمكن أيجاد التوزيعات التكرارية النسبية سواء فى حالة التوزيعات التكرارية البسيطة أو المزدوجة أو المتجمعة ، وسواء كانت المتغيرات كمية أو وصفية . وتعتبر التوزيعات النسبية على جانب كبير من الأهمية فى الاحصاء التحليلى إذ أنها تفيد فى حساب الاحتمالات التجريبية .

(٢-٣) العرض البيانى البيانات :-

ناقشنا فيما سبق الخطوط العريضة لمرحلتين من مراحل عملية التحليل الاحصائى وهما مرحلة جمع البيانات الاحصائية بنوعيها الحصر الشامل والعينة وأيضا مرحلة تبويب البيانات وتصنيفها (أى وضعها فى صورة جداول تمهيدا لعرضها وسهولة استيعابها) . والأن ننتقل لمرحلة أخرى من مراحل عملية التحليل الاحصائى وهى مرحلة عرض البيانات . ومما لا شك فيه أنه يمكن اعتبار

عملية تبويب البيانات فى صورة جداول بسيطة أو مركبة أو تكرارية بسيطة أو مزدوجة ، طريقة من طرق عرض البيانات وتسمى بالعرض الجدولى . غير أن اهتمامنا هنا ينصب على نوع أخر من طرق العرض وهو العرض البيانى للبيانات . وهناك عده طرق لعرض البيانات تتفق جميعها من حيث اعطاء فكرة صحيحة وسريعة عن كيفية تغيير الظواهر موضع الدراسة ولكنها تختلف عن بعضها البعض باختلاف البيانات المراد تمثيلها فهناك وسائل لعرض البيانات المبوبة (الموضوعة فى صورة فئات وتكرارات) نذكرها فيما بعد . أما أهم وسائل عرض البيانات غير المبوبة (الموضوعة فى صورتها الاولية الخام) فيمكن تلخيصها على النحو التالى :

(١) الرسوم البيانية

(٢) الأعمدة

(٣) الاشكال الهندسية (المستطيلات – الدوائر)

(٤) الصور

<u>أولا : الرسوم البيانية :-</u>

إذا أردنا مثلا تمثيل ظاهرتين بيانيا . فإننا نرسم محورين متعامدين يلتقيان فى نقطة تسمى الاصل . حيث يخصص أحد المحورين لتمثيل الظاهره الاولى والمحور الثانى لتمثيل الظاهره الثانية .

وإذا كانت احدى الظاهرتين تمثل الزمن فقد جرت العادة على تخصيص المحور الافقى (السينى) لمتثيلها . ثم نقسم كلا من المحورين - أكبر قيمة تمثل الظاهرة التى يراد تمثيلها بيانيا ، ثم ترصد نقطة الظواهر مبتدئين بتحديدها على المحور الافقى ، ثم نحدد موقعها على المحور الرأسى بحيث يكون بعدها عن المحور الافقى مساويا لقيمتها بالنسبة للتدرج الرأسى ، ثم نصل بين النقط المتتالية فنحصل على الخط البيانى الذى يمثل الظاهرة .

ومن المفضل دائما أن يكون رسم الخط البيانى قريبا من المحوريين لتسهيل قراءة ارقام التدريجين الممثلين لأى نقطة على الخط ولا يشترط تساوى التقسيم على المحورين لأن ذلك يتوقف على مقدار القيم وعددها كما انه لا يشترط تساوى مقياس الرسم على المحورين ، كذلك فإنه ليس من الضرورى بدء التدريج عند نقطة الأصل بل يفضل البدء بأصغر قيمة أو ما يقرب منهما إن كانت تبتعد كثيرا عن الصفر . وهذا ما يعبر عنه بـ (كسر المحاور) . وفى حالة ورود أكثر من ظاهرة موحده المقاييس فإن كل ظاهرة يمثلها خط بيانى على الرسم ويجب التمييز بين هذه الخطوط ، إما باستخدام الوان مختلفة يجعل إحداها خطا متصلا والأخر متقطعا أو على شكل نقط ... وهكذا . وأخيرا اذا تضمن الرسم البيانى عده ظواهر مختلفة المقاييس فيجب عندئذ رسم محاور اضافية رأسية بقدر تعدد المقاييس .

مثال (٦) :-

فيما يلى بيان باعداد الطلبة فى إحدى الكليات خلال السنوات ١٩٧٥-١٩٨٠ مقسمة حسب النوع :-

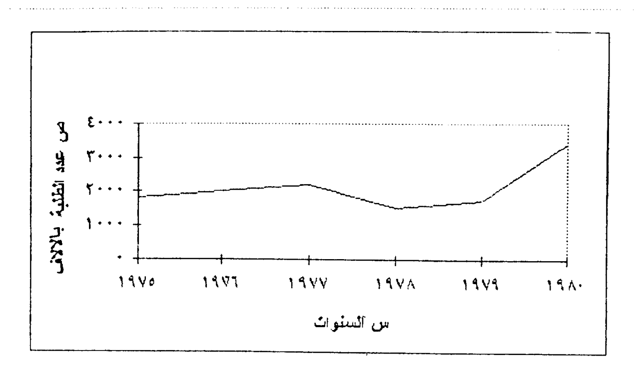
السنة	١٩٧٥	١٩٧٦	١٩٧٧	١٩٧٨	١٩٧٩	١٩٨٠
عدد الطلبة	١٨٠٠	٢٠٠٠	٢٢٠٠	١٥٠٠	١٧٥٠	٣٤٠٠
عدد الطالبات	١٢٠٠	١٥٠٠	١٠٠٠	٢١٠٠	١٤٠٠	٢٢٠٠
المجموع	٣٠٠٠	٣٥٠٠	٣٢٠٠	٣٦٠٠	٣١٥٠	٥٦٠٠

والمطلوب :-

- تمثيل اعداد الطلبة خلال السنوات ١٩٧٥-١٩٨٠ بخط بيانى
- تمثيل اعداد الطلبة والطالبات خلال السنوات المذكورة بخطوط بيانيه

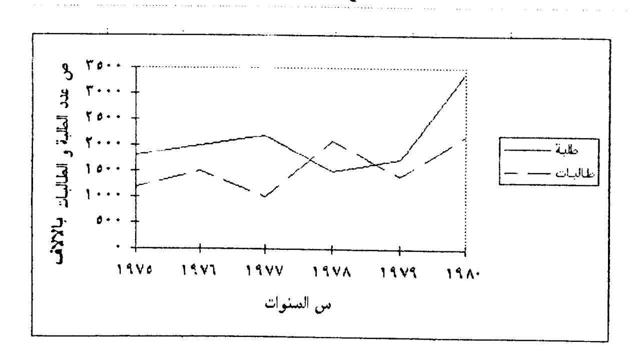
الحل
الشكلان الآتيان يوضحان الخطوط البيانية المطلوبة .

شكل (١)



• تمثيل اعداد الطلبة خلال السنوات(١٩٧٥-١٩٨٠) بيانيا

شكل ٢



• تمثيل اعداد الطلبة و الطالبات معا خلال السنوات(١٩٧٥-١٩٨٠) بيانيا

مثال (٧) :-

فيما يلى الكميات المنتجة بالألف الوحدات والمبيعات بالألف الجنيهات لاحدى السلع خلال الفترة من ١٩٧٠-١٩٨٠ .

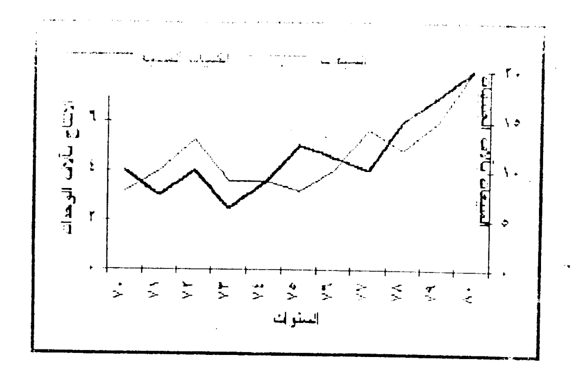
السنة	٧٠	٧١	٧٢	٧٣	٧٤	٧٥	٧٦	٧٧	٧٨	٧٩	٨٠
الكميات المنتجة	٤	٣	٤	٢,٥	٣,٥	٥	٤,٥	٤	٦	٧	٨
المبيعات	٨	١٠	١٣	٩	٩	٨	١٠	١٤	١٢	١٥	٢٠

والمطلوب :-

تمثيل الكميات المنتجة والمبيعات خلال الفترة (٧٠ - ٨٠) بخطوط بيانية .

الحل

حيث أن الكميات المنتجة تقاس بالوحدات والمبيعات مقاسة بالجنيهات فإن الرسم البيانى يظهر كما يلى :-



شكل توضيحى : الكميات المنتجه و المبيعات

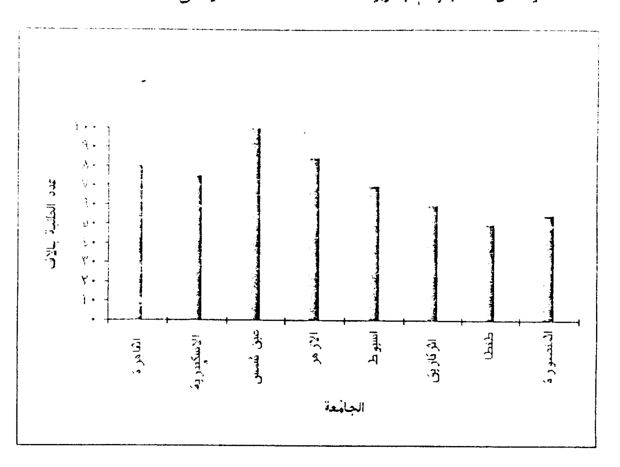
ثانيا : الأعمدة :-

تعتمد هذه الطريقة على اقامة عمود فوق كل نقطة علـى الاحداثـى السـينى يتناسب طوله مع قيمة الظاهرة .

فمثلا إذا كانت لدينا بيان بـاعداد الطلبـة بـالإلاف فـى بعض الجامعـات فـى سنة من السنوات :-

الجامعة	عدد الطلبة بالالاف
القاهرة	٨٠
الأسكندرية	٧٥
عين شمس	١٠٠
الأزهر	٨٥
أسيوط	٧٠
الزقازيق	٦٠
طنطا	٥٠
المنصورة	٥٥

فانه يمكن تمثيلهم بطريقة الأعمدة كما يلى : -

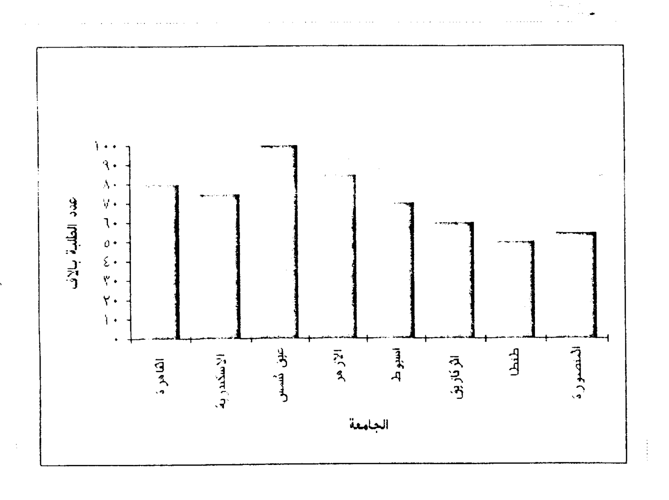


شكل توضيحى : الاعمده

ثالثا : الأشكال الهندسية :-

إذا كان لدى الباحث بعض الأرقام التى تمثل تغاير الظاهرة فى عدد قليل من السنين فإنه يفضل فى هذه الحالات تمثيلها بأشكال هندسية كالمستطيل أو الدائرة بحيث يكون التغير فى مساحة الشكل الهندسى متناسبا مع التغاير فى قيمة الظاهرة والطريقة الأكثر شيوعا هى استخدام المستطيلات الرأسية والتى تكون قواعدها متساوية وأطوالها متناسبة مع القيم التى تتخذها الظاهرة فى السنوات المختلفة .

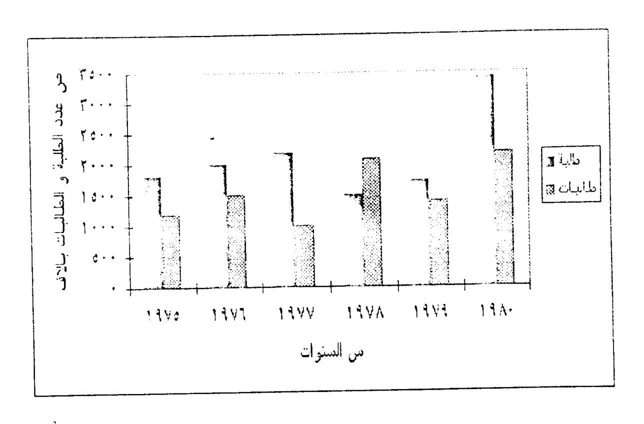
وعلى ذلك فإذا كانت لدينا البيانات السابقة الخاصة بأعداد الطلاب فى عدد من الجامعات فإن التمثيل باستخدام المستطيلات الرأسية يكون كما يلى :-



شكل (٥)

أما إذا أريد تمثيل أكثر من ظاهرة معينة بنفس الوحدات فإن الأمر يتطلب رسم مستطيلين أو أكثر لكل نقطة على الاحداثى السينى مع التمييز بين المستطيلات بألوان مختلفة منعا من الالتباس .

فإذا أردنا تمثيل بيانات مثال (٦) الخاصة باعداد الطلبة والطالبات فى إحدى الكليات فى الفترة من (١٩٧٥ - ١٩٨٠) فإن الرسم يكون كما يلى :-



شكل (٦)

وفى بعض الحالات يكون من الأنسب استخدام الدوائر بدلا من المستطيلات خاصة إذا كانت البيانات المطلوب عرضها خاصة بسنة أو سنين وكانت الظاهرة مقسمة إلى أجزاء . ففى هذه الحالة يمكن رسم دوائر تتناسب مساحتها مع قيم الظواهر ثم نقسم كل دائرة إلى قطاعات تتجمع فى مركز هذه الدوائر وتكون زوايا القطاعات متقاسمة مع قيم أجزاء الظاهرة .

ولتقسيم الدائرة فإننا ننسب أجزاء الظاهرة إلى مجموعها الكلى ثم نضرب كل نسبة فى ٣٦٠ (مقدار الزاوية عند مركز الدائرة) فتحصل على مقادير الزوايا المطلوبة . و فيما يلى بعض الامثلة التوضيحية :

مثال (٨) :-

الجدول الآتى يبين أعداد الطلاب سنة ١٩٨٠ حسب تقديراتهم آخر العام :

التقدير	عدد الطلبة
ضعيف جدا	١٥٠
ضعيف	٣٠٠
مقبول	٧٥٠
جيد	١٦٥٠
جيد جدا	١٢٠
ممتاز	٣٠

والمطلوب :-

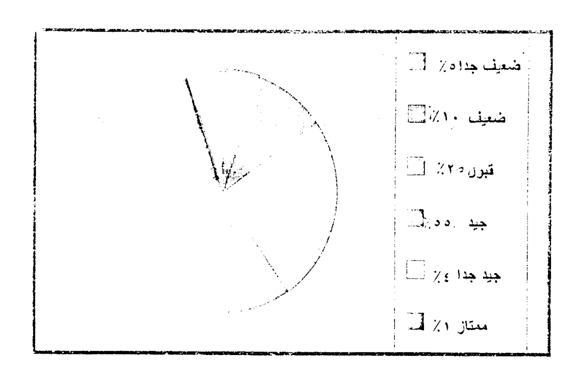
عرض البيانات السابقة باستخدام طريقة الدوائر

الحـــل

الجدول الآتى يبين خطوات الحل :-

التقدير	عدد الطلاب	النسبة	درجات الزاوية
ضعيف جدا	١٥٠	٠,٠٥	١٨
ضعيف	٣٠٠	٠,١	٣٦
مقبول	٧٥٠	٠,٢٥	٩٠
جيد	١٦٥٠	٠,٥٥	١٩٨
جيد جدا	١٢٠	٠,٠٤	١٤,٤
ممتاز	٣٠	٠,٠١	٣,٦
المجموع	٣٠٠٠	١	٣٦٠

وبذلك يكون شكل الدائرة كما يلى :-



شكل (٧) : التمثيل البيانى بأستخدام الدائره

وإذا كانت البيانات المطلوب عرضها خاصة بسنتين أو أكثر فإنه يجب استخدام دائرتين أو أكثر بحيث تكون النسب بين مساحات الدوائر متناسبة مع قيم الظواهر . وعلى سبيل المثال إذا كان العدد الكلى للطلبة سنة ١٩٨٠ = ٣٠٠٠ (المثال السابق) وكان عدد الطلبة فى سنة ١٩٨١ = ٤٥٠٠ فإن :-

$$\frac{\text{مساحة الدائرة الأولى}}{\text{مساحة الدائرة الثانية}} = \frac{٣٠٠٠}{٤٥٠٠}$$

$$\frac{[illegible]}{ط\ نق^2} = \frac{[illegible]}{٤٥} \quad ، \quad \frac{نق^2}{نق_2^2} = \frac{[illegible]}{[illegible]}$$

وعلى ذلك فإذا اعتبرنا نصف قطر الدائرة الأولى يساوى ١ سم فإن نصف قطر الدائرة الثانية يحدد كما يلى :-

$$\frac{١ \times ١}{نق_2^2} = \frac{٣٠}{٤٥}$$

$$نق_2^2 = \frac{٤٥}{٣٠}$$

نق = ١,٢٢ سم

وبذلك يمكن رسم كل من الدائرتين (الأولى نصف قطرها ١سم ، والثانية نصف قطرها ١,٢٢سم) ثم نقسم كل دائرة إلى قطاعات كما سبق .

رابعاً : طريقة الصور :-

فى هذا النوع نستخدم رموز أو صور معينة لها دلالة خاصة ، ذات صلة بموضوع الرسم . وهذه الطريقة يمكن للشخص العادى أن يفهمها بسرعة وسهولة ويستوعبها الفرد دون عناء . ومثال ذلك رسم طائرات ذات أحجام مختلفة للتعبير عن عدد الطائرات المنتجة فى السنوات المختلفة ، أو رسم أجهزة تليفزيون ذات أحجام مختلفة للتعبير عن انتاج أحد المصانع خلال فترة زمنية معينة .

[illegible] :-

تليفزيون ذات أحجام مختلفة للتعبير عن انتاج أحد المصانع خلال فترة زمنية معينة .

*** التمثيل البياني للتوزيعات التكرارية :-**

اما فـى حالة الجداول التكرارية (أو المتغيرات المتصلة) فيمكن التعبير عنها بيانيا بإحدى الصور الأربعة الاتية :

(١) المدرج التكرارى .

(٢) المضلع التكرارى .

(٣) المنحنى التكرارى .

(٤) المنحنى التكرارى المتجمع (الصاعد والهابط) .

*** المدرج التكرارى *Histogram* :-**

هو اسلوب لعرض البيانات الكمية (المتصلة) المبوبة باستخدام الاعمدة (المستطيلات) التى تتناسب مساحتها مع التكرارات وتتناسب قواعدها مع أطوال الفئات ، ولرسم مدرج تكرارى يمثل توزيعا تكراريا من هذا النوع تتبع الخطوات التالية :-

١. نرسم محورين متعامدين يخصص الأفقى منهما للفئات ، لذلك فليس من الضرورى أن يبدأ تدريجه من الصفر ولكن من فئة سابقة لأدنى فئات التوزيع مع مراعاة ترك طول فئة واحدة أو أكثر بعد أعلى فئات التوزيع . وأما المحور

الرأسى فيخصص للتكرارات ، لذلك فلابد وأن يبدأ تدريجه من الصفر وأن يراعى فى اختيار مقياس الرسم أن يكون مناسبا بحيث يسمح بإظهار أعلى تكرارات التوزيع .

٢. تقام مستطيلات متلاصقة قواعدها على المحور الأفقى وقواعد هذه المستطيلات هى أطوال الفئات المختلفة وارتفاعها هى التكرارات المناظرة لكل فئة . ولما كانت بيانات المتساوية الطول تمثلها أطوال متساوية على المحور الأفقى . لذلك فإن مساحة المستطيل الذى يمثل كل فئة تتحدد على اساس ارتفاع ذلك المستطيل أى على اساس التكرار المناظر لكل فئة . ولذلك تتناسب مساحات المستطيلات مع التكرارات (المساحة ما هى الا طول × عرض) . وبالتالى تكون مساحات تلك المستطيلات تساوى فى مجموعها المجموع الكلى للتكرارات لذلك فلابد وأن يكون التوزيع مقفلا وإلا اضطررنا الى أهمال الفئات المفتوحة ، وتكون مساحة المستطيلات المتبقية تقل فى مجموعها عن المجموع الكلى للتكرارات بما يناظر مجموع تكرارات الفئات المفتوحة المستبعدة من الرسم . وعليه فانه إذا استوجبت الضرورة استبعاد فئة أو أكثر من الرسم بسبب عدم تحديد نهاية أو بداية لها فلابد وأن يشار الى ذلك أسفل الرسم .

مثال (٩)

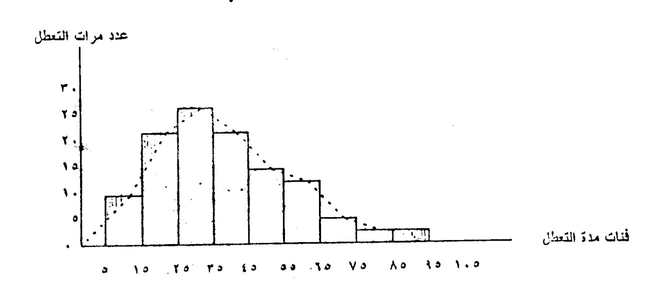
الجدول التالى يبين عدد مرات تعطل آلة فى مائـة مـره متتاليـة مقربـا إلـى أقرب دقيقة والمطلوب التعبير عن هذا التوزيع بيانيا .

جدول التوزيع التكرارى لمدد تعطل آلة فى مائة مرة متتالية

فئات تعطل الآلة بالدقيقة	-٥	-١٥	-٢٥	-٣٥	-٤٥	-٥٥	-٦٥	-٧٥	٨٥-٩٥
مراكز الفئات (س)	١٠	٢٠	٣٠	٤٠	٥٠	٦٠	٧٠	٨٠	٩٠
عدد مرات تعطل الآله او التكرارات (ك)	٨	١٢	٢٥	٢٠	١٣	٨	٢	١	١

الحـــل

المدرج و المضلع التكرارى لبيانات الجدول



قبل أن نبدأ دراستنا لشكل المضلع التكرارى لابد أولا من مراجعة تعريف مركز الفئة (س) .

مركز الفئة (س) : هى النقطة الوسطى التى تقع على بعدين متساويين من بداية الفئة ونهايتها . ولتعيين مركز الفئة نتبع أى من القواعد التالية :

مركز الفئة = (الحد الادنى للفئة + الحد الاعلى للفئة) / ٢

= الحد الادنى للفئة + نصف طول الفئة

= الحد الاعلى للفئة − نصف طول الفئة

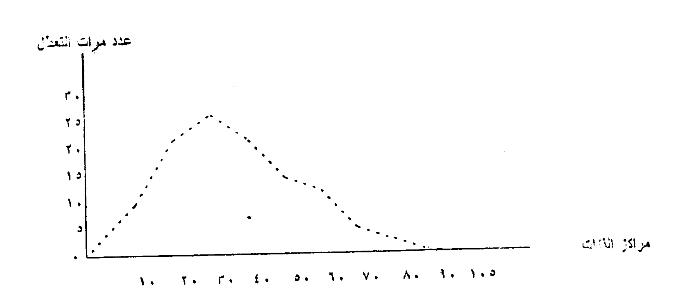
ويديهى أن الأمر أسهل من ذلك بكثير فى حالة الفئات متساوية الاطوال وإذ نكتفى بتطبيق أى من القواعد سالفة الذكر لتحديد مركز الفئة الاولى وتتحدد مراكز الفئات التالية بإضافة طول فئة الى مركز الفئة السابقة له مباشرة وهكذا حتى نهاية التوزيع .

والمضلع التكرارى ما هو الا خط بيانى منكسر يصل بين النقط التى احداثياتها مراكز الفئات (س ر) والتكرارات (ك ر) . وعلى ذلك فلتمثيل الجدول التكرارى بمضلع تكرارى نبدأ برسم المحوريين المتعامدين يخصص المحور الرأسى للتكرارات (ك ر) ، أما المحور الافقى فترصد عليه مراكز الفئات (س ر) . ثم

مستقيمة فإننا نحصل على مايسمى بالمضلع التكرارى .

وعلى ذلك فاذا أردنا التعبير عن بيانات مثال (٩) والخاص بتوزيع مدد تعطل الآلة بمضلع تكرارى كان الشكل كما يلى :

المضلع التكرارى لتوزيع مدد تعطل الآله



وأيضا يمكننا الحصول على المضلع التكرارى من المدرج التكرارى كما أوضحنا فى الرسم – وذلك بتصنيف القواعد العليا للمستطيلات وتوصيل هذه المنتصفات بخطوط مستقيمة مع إقفال الشكل عند بدايته ونهايته بنفس الطريقة المرضحة سابقا وذلك لأن منتصفات القواعد العليا للمستطيلات ما هى إلا مراكز الفئات .

* <u>المنحنى التكرارى *Frequency Curve*</u> :-

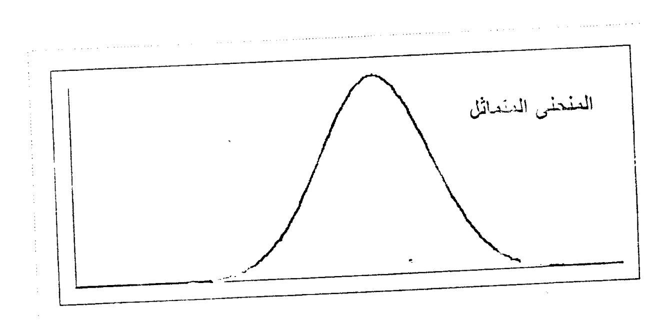
يمكن الحصول على المنحنى التكرارى من المضلع التكرارى عن طريق تمهيد جميع نقاط المضلع باليد أو بأى طريق رياضى آخر . وعملية التمهيد باليد هذه يمكن ان تختلف من شخص لآخر وحيث أنها عملية تقريبية فان المساحة تحت المنحنى تكون تقريبا مساويه للمساحة تحت المضلع وتحت المدرج التكرارى أيضا .

ومما يجب التنوية اليه أنه توجد أشكال مختلفة للمنحنى التكرارى ، حيث أنه لا يأخذ شكلا ثابتا الا انه يختلف من مجموعة الى أخرى ولكن هناك بعض المنحنيات الشائعة سنحاول الآن عرضها على سبيل المثال لا الحصر ومن خلالها سنتعرف على بعض التعريفات الهامة فى علم الاحصاء الوصفى ويمكن تقسيم هذه المنحنيات الى المجموعات الاتية :

• <u>**المنحنيات المتماثلة** *Symmetric Curves*</u>

يعرف المنحنى المتماثل بأنه المنحنى الذى لو اسقط من قمته عمودا لقسم المساحة تحت المنحنى الى جزئين متكافئين ومتماثلين كما هو واضح فى شكل (٤) وهو منحنى ذو نهاية عظمى فى منتصفه ثم يقترب من المحور الافقى تدريجيا على كل من جانبى هذه النهاية تقاربا متساويا من الجانبين . ويعتبر هذا

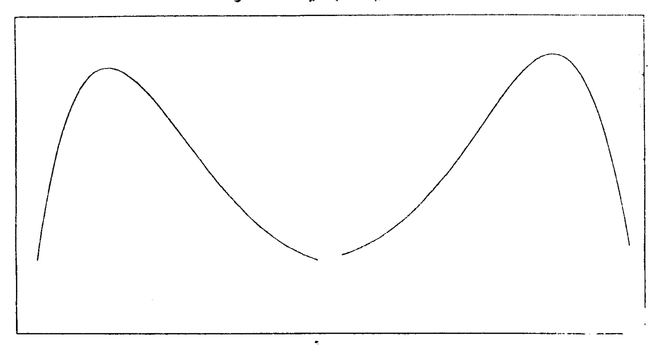
المنحنى من اشهر اشكال المنحنيات فهو يمثل الشكل المثال النظرى الذى نتوقع أن نحصل علية بدراسة عينه مستوفاه للشروط وعدد مفرداتها كبيرا جدا .



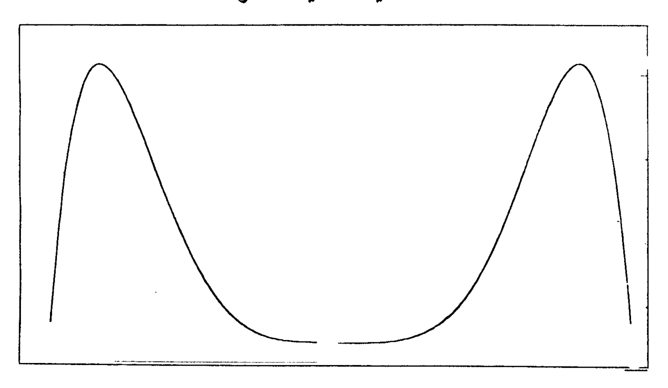
• المنحنيات الملتوية *Skewed Curves*.

أما اذا ابتعد المنحنى الممثل للظاهرة عن التماثل فاننا نطلق عليه منحنى ملتوى . وعادة ما يطلق على عدم تماثل المنحنى بالالتواء *Skewness* وقديكون هذا الالتواء بسيطا وقد يكون شديدا كما هو موضح فى الاشكال التالية :-

• منحنيات بسيطة الالتواء



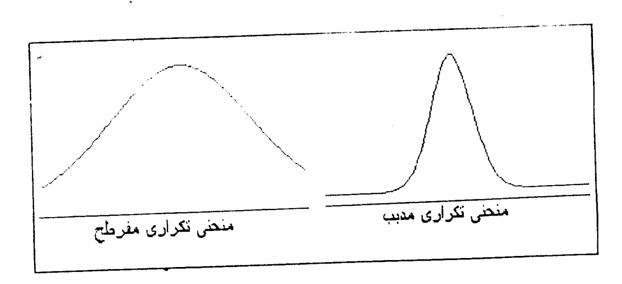
• منحنيات شديدة الالتواء



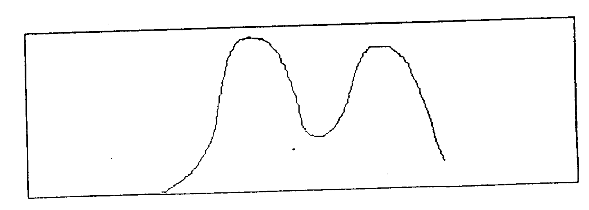
• المنحنيات المفرطحة والمدببة *Playkurtic & Leptokurtic Curves*

قد نحصل على منحنى متماثل ولكنه أكثر ضيقا فى الوسط من المنحنى الطبيعى (من خصائص المنحنى الطبيعى انه منحنى متماثل الالتواء ومعتدل

المتفرطح) وقمته أضيق وأكثر ارتفاعا من قمة المنحنى الطبيعى ويطلق على هذا المنحنى بأنه مدبب *Leptokurtic* أو قد يكون أكثر اتساعا فى وسطه من المنحنى المعتدل وقمته أيضا أكثر اتساعا من المنحى المعتدل ويطلق على المنحنى فى هذه الحالة بأنه منحنى متفرطح *Playkuartic* ويسمى هذا الاختلاف بنوعية بالتفرطح *Kurtosis* والشكل التالى يوضح حالتى المنحنى المدبب والمفرطح .



وأحيانا قد نحصل على منحنيات لها اكثر من قمة وذلك يأتى فى حالة عدم تجانس مفردات المجتمع ويطلق على هذا النوع من المنحنيات بالمنحنى متعدد القمم كما هو موضح فى الشكل التالى :



منحنى تكرارى متعدد القمم

ويشير هذا المنحنى الى أن البيانات موضع الدراسة مكونة من مجموعتين غير متجانستين وأول مثال على ذلك عند تمثيل توزيع عدد من المجرمين تبعا لمستويات ذكائهم فنجد أن متوسطى الذكاء بينهم أقلية بينما نجد غالبيتهم موزعة بين مرتفعى الذكاء ومنخفضى الذكاء .

*** المنحنى التكرارى المتجمع :-**

يعد المنحنى التكرارى المتجمع (الصاعد - الهابط) هو الطريقة الرابعة من طرق تمثيل بيانات الجداول التكرارية بيانيا . ويتم رسم المنحنى التكرارى المتجمع من واقع بيانات جدول التكرار المتجمع والذى سبق ذكره .

وحيث أنه يوجد لدينا نوعين من الجداول التكرارية المتجمعة وعليه فانه يوجد تبعا لذلك نوعين أيضا من المنحنيات التكرارية (منحنى متجمع صاعد - منحنى متجمع هابط) .

ولرسم المنحنى التكرارى المتجمع نبدأ برسم المحورين المتعامدين الافقى والرأسى بحيث يخصص المحور الأفقى للفئات أما المحور الراسى فيخصص للتكرارات المتجمعة امام نهايات الفئات (فى حالة التكرار المتجمع الصاعد) أو أمام بدايات الفئات (فى حالة التكرار المتمع الهابط) . كما يمكن رسم كل من المنحنيين المتجمعين الصاعد والهابط فى شكل واحد وذلك لأن نقطة التقاء (تقاطع) هذين المنحنيين يمكن من خلالها تحديد قيمة الوسيط بيانيا كما سنرى

فيما بعد . ومما ينبغى ملاحظته ان رسم اى من المنحنيين السابقين لا يتأثر بإنتظام أو عدم انتظام الفئات .

والمثال التالى يوضح كيفية رسم المنحنى التكرارى المتجمع الصاعد والهابط للجداول التكرارية .

مثال (١٠) :-

الجدول التكرارى التالى يبين توزيع درجات ٤٠ طالب فى احدى المواد :-

فئات	١٠-	١٥-	٢٠-	٢٥-	٣٠-	٣٥-	٤٠-	٤٥-٥٠	المجموع
تكرارات	٢	٦	٦	١١	٥	٤	٣	٣	٤٠

المطلوب :

١.رسم المنحنى التكرارى المتجمع الصاعد ثم أوجد عدد الطلبة الحاصلين على أقل من ٣٠ درجة .

٢.رسم المنحنى المتجمع الهابط ثم أوجد عدد الطلبة الحاصلين على ٢٠ درجة فأكثر .

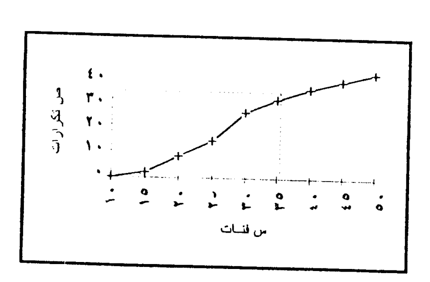
٣.رسم كل من المنحنيين الصاعد والهابط فى شكل ثم أوجد قيمة الوسيط بيانيا .

الحل

لرسم المنحنيين التكراريين الصاعد والهابط نكون أولا الجداول التكرارية المتجمعة الخاصة بكل منهما كما يلى :

الجدول التكرارى الأصلى		الجدول التكرارى المتجمع الصاعد		الجدول التكرارى المتجمع الهابط	
فئات	تكرارات	حدود عليا	تكرار متجمع صاعد	حدود دنيا	تكرار متجمع هابط
١٠-	٢	أقل من ١٠	صفر	١٠ فأكثر	٤٠
١٥-	٦	أقل من ١٥	٢	١٥ فأكثر	٣٨
٢٠-	٦	أقل من ٢٠	٨	٢٠ فأكثر	٣٢
٢٥-	١١	أقل من ٢٥	١٤	٢٥ فأكثر	٢٦
٣٠-	٥	اقل من ٣٠	٢٥	٣٠ فأكثر	١٥
٣٥-	٤	أقل من ٣٥	٣٠	٣٥ فأكثر	١٠
٤٠-	٣	أقل من ٤٠	٣٤	٤٠ فأكثر	٦
٤٥-٥٠	٣	أقل من٤٥	٣٧	٤٥ فأكثر	٣
		اقل من ٥٠	٤٠	٥٠ فأكثر	صفر
المجموع	٤٠				

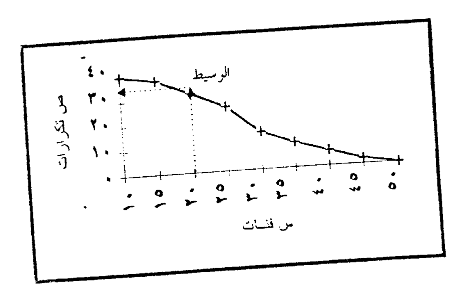
(١) المنحني التكراري المتجمع الصاعد



شكل (١٦)

التعليق : من الشكل نجد أن عدد الطلاب الحاصلين على اقل من ٣٥ درجـة يعـادل ٣٠ طالب

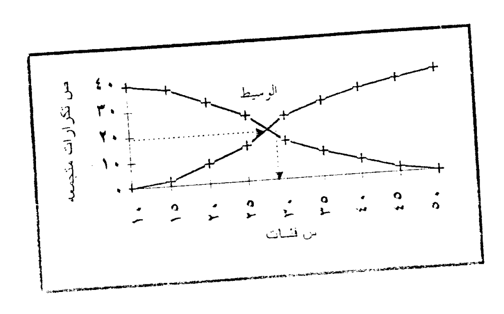
(٢) المنحنى التكرارى المتجمع الهابط :



شكل (١٧)

من الشكل السابق نجد أن عدد الطلاب الحاصلين على ٢٠ درجة فأكثر يعادل ٣٢ طالبا .

(٣) المنحنيان التكراريان التجميعيان الصاعد والهابط معا



شكل (١٨)

الباب الثالث

مقاييس النزعة المركزية

Measures of Central Tendancy

ناقشنا فيما سبق كيفية استخدام الجداول الاحصائية والخرائط البيانية للوصول الى قياس وصفى للبيانات الاحصائية ، فى هذا الباب والابواب التالية سوف نستكمل الصورة بوصف وتلخيص كيفية توزيع هذه البيانات فى صورة رقمية مستخدمين بعض المقاييس الاحصائية لقياس القيمة المتوسطة التى تتركز حولها البيانات – أى المتوسطات – وكذلك لقياس درجة انتشار المشاهدات او تركزها – أى مقاييس التشتت أو الانتشار ، علاوة على التعرض بشىء من التفصيل الى خصائص هذه التوزيعات التكرارية والمقارنة بينها ، وذلك من خلال مناقشة مقاييس احصائية أخرى تقيس مدى تماثل التوزيع أو التوائه وتقيس كذلك درجة تفرطح التوزيع أو تحدبه .

كما نود أن نشير الى أنه لمعظم أن لم يكن لكل هذه المقاييس فقد اعدت برامج لحسابها باستخدام الحاسبات الالية مما يغنينا من عناء حسابها حين نحتاج اليها ولكن يظل حدود واستخدامات كل مقياس منها مسئولية الفرد المستخدم له وليس مسئولية الحاسب .

فى هذا الباب سوف نتناول المقصود بالنزعة المركزية ونتعرف على أهم المقاييس الاحصائية المستخدمة لقياسها وذلك فى حالتى البيانات الغير مبوبة والبيانات المبوبة .

يعنى المتوسط (*Average*) لأى مجموعة من القراءات أنه القيمة التى تتوسطها وتتركز عندها - أو قريبا منها - ولذلك تعرف أحيانا هذه المقاييس بمقاييس النزعة المركزية . وتوجد عدة مقاييس للنزعة المركزية وهذا أمر طبيعى حيث أن البيانات تختلف فى طبيعتها فمثلا وجدنا فى الباب السابق توزيعات متماثلة وتوزيعات بسيطة الالتواء وتوزيعات شديدة الالتواء وأخرى مفرطحة .... الخ .

ومن هنا كانت الحاجة إلى وجود أكثر من مقياس للنزعة المركزية وعلينا نختار المقياس الذى يناسب التوزيع التكرارى موضوع الدراسة ، فمن هذه المقاييس من يتأثر بالقراءات الشاذة لدرجة أنه لا يعبر عن المجموعة أصدق تعبير .

(٤-١) الوسط الحسابى *Arithmatic Mean* :-

يعتبر الوسط الحسابى من أكثر مقاييس النزعة المركزية استخداما حيث أنه سهل الفهم وبسيط فى الحساب ويصلح للمقارنة بين المجموعات المختلفة ويعرف بأنه عبارة عن مجموع القيم مقسوما على عددها .

(٢-١-١) حساب الوسط الحسابي من البيانات غير المبوبة :-

يمكن حساب الوسط الحسابي لمجموعة من بيانات ظاهرة ما بعدة طرق تختلف فيما بينها من حيث الجهد الحسابي المبذول ولكنها تؤدى في النهاية الى الحصول على نفس النتيجة كما سنرى من التحليل التالي :

أولاً : الطريقة العادية :-

بفرض أنه لدينا مجموعة من مفردات ظاهرة ما تأخذ القيمة الاتية :

$\text{س}_1 ، \text{س}_2 ، \text{س}_3 ، \cdots\cdots ، \text{س}_\text{ن}$. حيث (ن) تمثل عدد قيم المتغير فان الوسط الحسابي والذى نرمز له بالرمز ($\bar{\text{س}}$) يتم حسابه كما يلى :

$$\bar{\text{س}} = \frac{\text{مجموع القيم}}{\text{عدد القيم}} = \frac{\text{مجـ س}_\text{ر}}{\text{ن}} = \frac{\text{س}_1 + \text{س}_2 + \cdots\cdots + \text{س}_\text{ن}}{\text{ن}}$$

وللتسهيل يمكننا كتابة هذه الصيغة على النحو التالي :

$$\bar{\text{س}} = \frac{\sum_{\text{ر}=1}^{\text{ن}} \text{س}_\text{ر}}{\text{ن}}$$

حيث تشير مجـ س الى مجموع قيم المتغير (س) والتى عددها (ن) مفردة .

فإذا كان لدينا القيم ٢٧ ، ١٨ ، ٣٥ ، ١٩ ، ٣١ فإن الوسط الحسابي

يصبح :-

$$\bar{س} = \frac{\sum_{ر=1}^{ه} س_ر}{ن} = \frac{٢٧ + ١٨ + ٣٥ + ١٩ + ٣١}{٥} = ٢٦$$

وللوسط الحسابي مجموعة من الخصائص التي تميزه عن غيره من المتوسطات الأخرى والتي سوف يتم الاستفادة منها في التعرف على الصورة الأخرى لحساب الوسط الحسابي .

[illegible]

[illegible] في حساب الوسط الحسابي [illegible] ذات الظاهرة التي تقضي بأن الوسط الحسابي يتأثر بعمليتي الجمع والطرح .

فمثلا اذا كان لدينا القيم $س_1$ ، $س_2$ ، ٠٠٠٠ ، $س_ن$ وطرحنا ثابت (أ)

من كل مفرده لحصلنا على النتائج التالية :

$$(س_1 - أ) ، (س_2 - أ) ، .... (س_ن - أ)$$

فإن الوسط الحسابي المحسوب لهذه القيم الجديدة يأخذ الصورة الآتية :

$$\bar{س} = \frac{\sum (س_ر - أ)}{ن} + أ \quad .... (١)$$

وايضا لو تم جمع المقدار الثابت (أ) على كل مفرده من مفردات العينـة فان النتائج سوف تكون على النحو التالى :-

$$(\text{س}_1 + \text{أ})\ ،\ (\text{س}_2 + \text{أ})\ ،\ \ldots\ldots\ldots\ (\text{س}_\text{ن} + \text{أ})$$

ويكون الوسط الحسابى فى الصورة

$$\bar{\text{س}} = \frac{\text{مجـ}\ (\text{س}_\text{ر} + \text{أ})}{\text{ن}} - \text{أ} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ (٢)$$

وتؤدى طريقة الفروق البسيطة الى اختصار العمليات الحسابية وخصوصا اذا كانت المفردة كبيرا مما يؤدى الى تخفيض واضح فى الجهد الحسابى وايضا فى احتمالات الخطأ .

وتعتمد طريقة الفروق البسيطة على الشكل الرياضى للمعادلة (١) إلا أنه يمكننا اختصارها إلى الصورة التالية .

$$\bar{\text{س}} = \text{أ} + \frac{\text{مجـ ح}_\text{ر}}{\text{ن}}$$

حيث أ أى وسط فرضى من مجموعة القيم . $\text{ح}_\text{ر} = (\text{س}_\text{ر} - \text{أ})$ ،

$\text{ر} = ١ ، ٢ ، ٣ ، \ldots\ldots\ldots\ ، \text{ن}$

مثال (٢) :-

البيانات الاتية تمثل الربح السنوى بالجنيه لاحدى الشركات فى الخمس سنوات الاخيرة :

[illegible]	[illegible]	[illegible]	[illegible]	١٩٩٤	١٩٩٥
الربح	٣٢٠٠	٣٦٢٠	٤١٤٠	٢٩٩٠	٣١٥٠

المطلوب : حساب متوسط الربح السنوى للشركة خلال المدة السابقة وذلك بالطريقتين (العادية – الفروق البسيطة) .

الحل :

للحصول على متوسط الربح السنوى بالطريقتين (العاديـة – الفـروق البسيطة) لابد أولا من تكوين الجدول التالى :

$س_ر$	$ح_ر$ = ($س_ر$ – أ)
٣٢٠٠	٣٢٠٠ – ٣٦٢٠ = –٤٢٠
٣٦٢٠	٣٦٢٠ – ٣٦٢٠ = صفر
٤١٤٠	٤١٤٠ – ٣٦٢٠ = ٥٢٠
٢٩٩٠	٢٩٩٠ – ٣٦٢٠ = –٦٣٠
٣١٥٠	٣١٥٠ – ٣٦٢٠ = –٤٧٠
مجـ ١٧١٠٠	– ١٠٠٠

و الطريقة العادية :

$$\bar{س} = \frac{\sum_{ر=١}^{د} س_ر}{د}$$

$$= \frac{١٧١٠٠}{٥} = ٣٤٢٠ \text{ جنيه}$$

• طريقة الفروق البسيطة :

$$\overline{س} = أ + \frac{\sum_{ر=١}^{د} ح_{ر}}{ن}$$

$$= ٣٦٢٠ + \frac{-١٠٠٠}{٥} = ٣٤٢٠ \text{ جنيه}$$

يتضح لنا من المثال السابق أن متوسط ربح الشركة السنوى خلال الخمس سنوات الاخيرة يساوى ٣٤٢٠ جنيه . كما يتضح لنا انه اذا كانت قيمة الوسط الحسابى ($\overline{س}$) بالطريقة العادية فانه يمكن معرفة مجموع القيم وفقا للمعادلة الاتية :

$$\boxed{\sum_{ر=١}^{ن} س_{ر} = ن \, \overline{س}}$$

ففى المثال السابق نجد أن حاصل ضرب متوسط الربح السنوى الناتج باحدى الطريقتين (٣٤٢٠) فى عدد السنوات وهو (٥) يساوى ١٧١٠٠ جنيه والذى يمثل مجموع الارباح خلال الفترة كلها .

ثالثا : طريقة الفروق المعدلة :-

[illegible]

بأن الوسط الحسابى يتأثر بعمليتى الضرب والقسمة . أى أنه لو ضربنا مفردات العينة فى مقدار ثابت (ب) فان :

$$\bar{س} = \frac{ب س_1 + ب س_2 + \dots + ب س_ن}{ب ن}$$

والوسط الحسابى الجديد ($\bar{س}'$) لهذه القيم هو :

$$\bar{س} = \frac{مج ب س_ر}{ن} = \frac{ب مج س_ر}{ن} = ب \bar{س}$$

معنى ذلك أن الوسط الحسابى الجديد هو نفس الوسط الحسابى القديم مضروبا فى المقدار الثابت (ب) ، وللحصول على الوسط الحسابى القديم للبيانات نقوم بقسمة الوسط الجديد على المقدار الثابت السابق الضرب فيه (ب) (ونعامل القسمة نفس معاملة الضرب) .

وعموما تقتضى طريقة الفروق المعدلة بأنه اذا وجدنا أن جميع قيم الفروق البسيطة (ح) تقبل القسمة على مقدار ثابت (ب) بدون باقى فانه يمكن اختصار العمليات الحسابية أكثر عند حساب الوسط الحسابى كالاتى :

- للحصول على الفروق المعدلة (ح') نقسم جميع قيم الفروق البسيطة (ح) على مقدار ثابت (ب) أن امكن حيث :

$$ح' = \frac{ح}{ب}$$

ولإيجاد الوسط الحسابى للقيم الاصلية يجب أن نجرى تصحيحا للعمليات السابقة بحيث نضيف الوسط الفرضى (أ) والذى سبق طرحه من القيم ثم نضرب فى المقدار الثابت (ب) والذى سبق القسمة عليه أيضا وبالتالى يأخذ الوسط الحسابى للقيم الاصلية الصيغة الاتية :

$$\bar{س} = أ + \frac{مجـ\ ح'}{ن} \times ب$$

مثال ٢ :

احسب الوسط الحسابى لمجموعة القيم الواردة فى المثال (٢) .

<u>الحل :</u>

لاحظنا فى المثال أنه تم اختيار الوسط الفرضى أ = ٣٦٢٠ وذلك للحصول على الفروق البسيطة ($ح_ر$) . وللحصول على الفروق المعدلة ($ح'_ر$) نقسم جميع قيم ($ح_ر$) على الثابت (ب = ١٠) وبالتالى يمكننا حساب الوسط الحسابى باستخدام طريقة الفروق المعدلة على النحو التالى :

$س_ر$	$ح_ر$	$ح'_ر = \frac{ح_ر}{ب} = \frac{ح_ر}{١٠}$
٣٢٠٠	-٤٢٠	
٣٦٢٠	صفر	-٤٢
٤١٤٠	٥٢٠	صفر
٢٤١٠	-١?٠	٥٢
[illegible]	-٤٧٠	-١?
مجـ = ١٧١٠٠	[illegible]	-٤٧

$$\overline{\text{س}} = \text{أ} + \frac{\text{مجـ ح}}{\text{ن}} \times \text{ب}$$

$$= ٣٦٢٠ + \frac{١٠٠}{٥} \times ١٠ = ٣٤٢٠ \text{ جنيه}$$

وهى نفس النتيجة التى حصلنا عليها سواء بالطريقة العادية أو بطريقة الفروق البسيطة . ومما يجب ملاحظته أن هناك بعض الخصائص الاخرى للوسط الحسابى نذكر منها :

٢- مجموع انحرافات القيم عن وسطها الحسابى يساوى الصفر .

الاثبات

نفرض أن المتغير $\text{س}_\text{ر}$ يأخذ القيم $\text{س}_١$ ، $\text{س}_٢$ ، .... ، $\text{س}_\text{ن}$ بوسط حسابى ($\overline{\text{س}}$) . فتكون انحرافات (فروق) هذه القيم عن وسطها الحسابى هى ($\text{س}_١ - \overline{\text{س}}$) ، ($\text{س}_٢ - \overline{\text{س}}$) ، ..... ، ($\text{س}_\text{ن} - \overline{\text{س}}$) وبأخذ المجموع لهذه القيم نجد أن :

$$\sum_{\text{ر}=١}^{\text{ن}} (\text{س}_\text{ر} - \overline{\text{س}}) = \sum_{\text{ر}=١}^{\text{ن}} \text{س}_\text{ر} - \sum_{\text{ر}=١}^{\text{ن}} \overline{\text{س}}$$

$$= \text{ن}\,\overline{\text{س}} - \text{ن}\,\overline{\text{س}} = \text{صفر}$$

ويمكن الاعتماد على هذه الخاصية فى توضيح الدور المركزى الذى يلعبه الوسط الحسابى .

٤ – مجموع مربعات انحرافات القيم عن وسطها الحسابى أقل ما يمكن :

أى المطلوب اثبات أن

$$\sum_{ر=1}^{ن} (س_ر - \bar{س})^2 < \sum_{ر=1}^{ن} (س_ر - أ)^2$$

حيث (أ) هو أى مقدار ثابت بخلاف الوسط الحسابى ($\bar{س}$)

الاثبات

بالطرف الايسر للمتباينة السابقة نجد أن :

$$\sum_{ر=1}^{ن} (س_ر - أ)^2 = \sum_{ر=1}^{ن} [(س_ر - \bar{س}) + (\bar{س} - أ)]^2$$

$$= \sum_{ر=1}^{ن} [(س_ر - \bar{س})^2 + ٢ (س_ر - \bar{س}) (\bar{س} - أ) + (\bar{س} - أ)^2]$$

$$= \sum_{ر=1}^{ن} (س_ر - \bar{س})^2 + ٢ (\bar{س} - أ) \sum_{ر=1}^{ن} (س_ر - \bar{س}) + ن (\bar{س} - أ)^2$$

حيث $\bar{س}$ ، أ تمثل مقادير ثابتة وايضا وبالاعتماد على الخاصية السابقة التى تنص بأن مج (س – $\bar{س}$) = صفر فان :

$$مج (س_ر - أ)^2 = [\sum_{ر=1}^{ن} (س_ر - \bar{س})^2 + ن (\bar{س} - أ)^2]$$

(وحيث [illegible] ن (س – أ)² يظل مقدار [illegible]

[illegible] $)^2$ + [illegible]

وبالتالى نجد أن

$$\sum_{ر=1}^{ن} (س_ر - \bar{س})^{2} < \sum_{ر=1}^{ن} (س_ر - أ)^{2}$$

مثال (٣) :-

اليك مجموعة من القيم المأخوذة من سجلات احدى الشركات وكانت

كالتالى : ٦ ، ٨ ، ٩ ، ١٢ ، ١٠

المطلوب :

١-اثبت أن انحرافات هذه القيم عن وسطها الحسابى يساوى صفرا .

٢-أوجد المقادير التالية

$$\sum_{ر=1}^{٥} (س - ٩)^{2} \quad ، \quad \sum (س - ٧)^{2}$$

الحل :

نحسب أولا الوسط الحسابى للقيم $\bar{س}$:

$$\bar{س} = \frac{\sum س_ر}{ن} = \frac{٤٥}{٥} = ٩$$

ثم نكون الجدول التالى :-

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
٦	٦ – ٩ = –٣	٩	٦ – ٧ = –١	١
٨	= –١	١	= +١	١
٩	= صفر	صفر	= +٢	٤
١٢	= +٣	٩	= +٥	٢٥
١٠	= +١	١	= +٣	٩
مجـ = ٤٥	صفر	٢٠	– ٠	٤٠

مما سبق ، يلاحظ أن مجموع انحرافات القيم عن وسطها الحسابى يساوى صفر (الخاصية الثالثة) ، كما يلاحظ ايضا أن مجموع مربعات انحرافات القيم السابقة عن وسطها الحسابى (٩) يساوى ٢٠ فى حين أن مجموع مربعات انحرافات القيم عن اى وسط فرضى آخر كما فى المثال يساوى ٤٠ (الخاصية الرابعة) .

٥- الوسط الحسابى لمجموع (أو الفرق بين) أزواج القيم يساوى مجموع الوسطين الحسابين لكل مجموعة من القيم (أو الفرق بينهما) . فاذا كان لدينا متغيرين س ، ص وجمعنا قيمتهما (أو طرحناهما) فان الوسط الحسابى للمجموع (أو الفرق) يساوى مجموع الوسطين الحسابيين أو (الفرق بينهما) حيث :

$$\frac{\sum_{ر=1}^{ن}(س_ر \pm ص_ر)}{ن} = \frac{\sum_{ر=1}^{ن} س_ر \pm \sum_{ر=1}^{ن} ص_ر}{ن}$$

$$= \frac{\sum_{ر=1}^{ن} س_ر}{ن} \pm \frac{\sum_{ر=1}^{ن} ص_ر}{ن}$$

$$= \bar{س} + \bar{ص}$$

ومما يلزمنى ملاحظته أنه يمكن تعميم هذه الخاصيــة على أكثر مـن متغيرين .

٦- اذا قسمت مجموعة البيانات المكونة من (ن) مفرده الى قسمين يتكون الاول من ($ن_1$) مفرده والثانى من ($ن_2$) مفرده فـان الوسـط الحسـابى العـام لمجموعة القيم ككل هو عبارة عن الوسط الحسابى المرجح لمتوسطى المجموعتين والترجيحات هنا هى عدد المفردات لكل مجموعة على حده :

أى أن :

$$\bar{س} = \frac{ن_1 \bar{س}_1 + ن_2 \bar{س}_2}{ن_1 + ن_2}$$

وايضا تجدر الاشارة الى انه يمكننا تعميم هذه الخاصية على اكثر من مجموعتين من القيم . علاوة على امكانية الاستفادة من هذه الخاصيـة فى حساب الوسط الحسابى للتوزيعات التكرارية الغير منتظمة كما سنرى فيما بعد .

(٣-١-٢) الوسط الحسابى للبيانات المبوبة (بيانات موزعة [illegible]) :-

سبق أن أوضحنا أن البيانات قد تكون فى صورتها الاصلية (الاولية) أى لم تخضع لأى تلخيص (وهذا ما يعرف بالبيانات الغير مبوبة) أو يجرى عليها نوع من التلخيص بقصد سهولة عرضها واستخلاص النتائج منها وذلك بتقسيم البيانات الاصلية الى عدد من [illegible] (فئات) ويستخدام مركز الفئة [illegible] لكل القيم الموجودة داخل الفئة : بمعنى انه لو ضرب مركز الفئة فى عدد التكرارات المناظرة لكل فئة فان الناتج يعبر عن مجموع المفردات التى توجد داخل هذه الفئة بحيث أن مجموع انحرافات المفردات عن مركز الفئة يساوى صفر (الخاصية الثالثة) وفى بعض الاحيان قد لا يكون مركز الفئة ممثلا لمتوسط المفردات داخل الفئة مما ينتج عنه بعض التقريب فى الحساب ويكون ذلك واضح فى حالة التوزيعات غير المنتظمة حيث يكون الخطأ الناشىء عن استخدام مركز الفئة [illegible] عن متوسط مفردات الفئة قريبا من الصفر بحيث يكون الخطأ [illegible] .

وقياسا على ما سبق توضيحه فى حالة حساب الوسط الحسابى من بيانات غير مبوبة فاننا سوف نستعرض الثلاث طرق المعروفة لحساب الوسط الحسابى من التوزيعات التكرارية وهى :-

- الطريقة العادية
- طريقة الفروق البسيطة
- طريقة الفروق المعدلة

أولا : الطريقة العادية :-

تتلخص خطوات العمل طبقا لهذه الطريقة فى الاتى :

١- ايجاد مراكز الفئات (س ر) حيث

$$\text{مركز الفئة} = \frac{\text{الحد الادنى} + \text{الحد الاعلى}}{2}$$

٢- يتم ضرب مركز كل فئة (س ر) فى التكرار المناظر له (ك ر) وذلك للحصول على عمود س ر ك ر .

٣- يمكن الحصول على الوسط الحسابى ($\bar{س}$) عن طريق قسمة حاصل جمع العمود السابق (مجـ س ر ك ر) على مجموع التكرارات (مجـ ك ر) .

$$\bar{س} = \frac{\sum_{ر=1}^{ن} [illegible]}{\sum_{ر=1}^{ن} [illegible]}$$

حيث تعبر (ن) هنا عن عدد فئات جدول التكرار والتكرارات وليس عدد [illegible]

[illegible]

مثال (٢) :-

فيما يلي بيان الأجر اليومي لعدد من العمال بإحدى الشركات :

فئات الأجر اليومي	٢–	٤–	٦–	٨–	١٠–	١٢–١٤	المجموع
عدد العمال	٣	٥	٢٢	٤٠	[illegible]	١٠	١٠٠

المطلوب : حساب الوسط الحسابي بالطريقة العادية للأجر اليومي .

الحل :

طبقا لهذه الطريقة نقوم بحساب مراكز الفئات (س ر) أولا ثم نضرب كلا منها في التكرار المناظر (ك ر) ونحصل على (ك ر س ر) ونقسم مجموع التكرارات (مجـ ك ر) فنحصل على الوسط الحسابي وذلك كما في الجدول

[illegible]

فئات الاجر	عدد العمال ($ك_ر$)	مركز الفئة ($س_ر$)	$س_ر ك_ر$
٢-	٣	٣	٩
٤-	٥	٥	٢٥
٦-	٢٢	٧	١٥٤
٨-	٤٠	٩	٣٦٠
١٠-	٢٠	١١	٢٢٠
١٢-١٤	١٠	١٣	١٣٠
المجموع	١٠٠	——	٨٩٨

حيث $\sum_{ر=١}^{٦} ك_ر = ١٠٠$ ، $\sum_{ر=١}^{٦} س_ر ك_ر = ٨٩٨$

$$\therefore \overline{س} = \frac{\sum_{ر=١}^{٦} س_ر ك_ر}{\sum_{ر=١}^{٦} ك_ر} = \frac{٨٩٨}{١٠٠} = ٨,٩٨ \text{ جنيه}$$

ثانيا : طريقة الفروق البسيطة :-

طبقا لهذه الطريقة نحسب أولا مراكز الفئات ($س_ر$) ثم نختار أحدها كوسط فرضي (أ) (يفضل مركز الفئة المقابل لاكبر تكرار) .

- يطرح الوسط الفرضى الذى تم اختياره من كل مركز من مراكز الفئات نحصل على الفروق البسيطة ($ح_ر$) كما فى حالة البيانات غير المبوبة .

- يتم ضرب كل قيمة من قيم الفروق البسيطة ($ح_ر$) فى التكرارات المناظرة ($ك_ر$) ثم نجمع قيم العمود الناتج نحصل على القيمة :

$$\sum_{ر=1}^{ن} ح_ر \, ك_ر$$

- وبقسمة قيمة $\sum_{ر=1}^{ن} ح_ر \, ك_ر$ على مجموع التكرارات ($\sum ك_ر$)

و باضافة الوسط الفرضى (أ) الذى تم اختياره من بين قيم ($س_ر$) الى ناتج القسمة نحصل على الوسط الحسابى على ضوء العلاقة الآتية :

$$\bar{س} = أ + \frac{\sum ح_ر \, ك_ر}{\sum ك_ر}$$

مثال (٥)

فى مثال (٤) أوجد الوسط الحسابى بطريقة الفروق البسيطة .

<u>الحل</u>

الجدول التالى يبين خطوات الحل وذلك بأخذ القيمة ([illegible]) كوسط فرضى ثم نكون الجدول التالى :

[illegible]	التكرار (ك)	مركز [illegible] (س)	[illegible]	[illegible]
٢ -	٣	٣	٣-٩ = -٦	-١٨
٤ -	٥	٥	٥-٩ = -٤	-٢٠
[illegible]	٢٢	[illegible]	٧-٩ = -٢	[illegible]
[illegible]	٤٠	[illegible]	٩-٩ = صفر	[illegible]
[illegible]	٢٠	١١	١١-٩ = ٢	٤٠
[illegible]	١٠	١٠	١٣-٩ = ٤	٤٠
				[illegible]

[illegible]

[illegible] = [illegible] مجـ ك = ١٠٠ مجـ ح [illegible] = [illegible]

[illegible]

[illegible] = ١٨ [illegible]

[illegible]

نلاحظ في الحل السابق أن الفروق البسيطة (ح) والتي [illegible]

الفروق [illegible] قبل جميعها [illegible] على طول الفئة (ل = ٢) [illegible]

الفروق المعدلة (ح) على قسمة [illegible] الفروق (ح) على طول الفئة (ل = ٢) [illegible]

[illegible]

[illegible]

و بالضرب في العامل المشترك (طول الفئة ب) واضافة الناتج الى قيمة الوسط الفرضى (أ) كما مر واتضح فى تطبيق هذه الطريقة فى حالة البيانات الغير مبوبة مع أخذ التكرارات فى الاعتبار أى تصبح العلاقة على النحو التالى :

$$\overline{\text{س}} = \text{أ} + \frac{\text{مج } \text{ح}'_{\text{ر}} \text{ك}_{\text{ر}}}{\text{مج } \text{ك}_{\text{ر}}} \times \text{ب}$$

حيث : $\text{ح}'_{\text{ر}} = \frac{\text{ح}_{\text{ر}}}{\text{ب}}$ ، ب = طول الفئة

مثال ٦ :

فى مثال (٤) أوجد الوسط الحسابى مستخدما طريقة الفروق المعدلة.

الحل

الجدول التالى يبين خطوات الحل وذلك بأخذ القيمة (أ = ٩) كوسط فرضى وطول الفئة (ب = ٢) كعامل مشترك .

الفئات	التكرارات ($\text{ك}_{\text{ر}}$)	مراكز الفئات ($\text{س}_{\text{ر}}$)	$\text{ح}_{\text{ر}} = \text{س}_{\text{ر}} - ٩$	$\text{ح}'_{\text{ر}} = \frac{\text{ح}_{\text{ر}}}{٢}$	$\text{ح}'_{\text{ر}} \text{ك}_{\text{ر}}$
٢ -	٣	٣	٦-	٣-	٩-
٤ -	٥	٥	٤-	٢-	١٠-
٦ -	٢٢	٧	٢-	١-	٢٢-
٨ -	٤٠	٩	صفر	صفر	صفر
١٠ -	٢٠	١١	٢	١	٢٠
١٢ - ١٤	١٠	١٣	٤	٢	٢٠
المجموع	١٠٠	—	—	—	٤٠ - ٤١ = -١

وحيث أن : أ = ٩ ، ب = ٢ ، مجـ $\text{ك}_{\text{ر}}$ = ١٠٠ ، مجـ $\text{ح}'_{\text{ر}} \text{ك}_{\text{ر}}$ = -١

$$\bar{س} = أ + \frac{مج\ ح\ ك}{مج\ ك} \times ب$$

$$= ٩ + \frac{-١}{١٠٠} \times ٢ = ٨,٩٨$$ (نفس النتيجة السابقة)

مثال (٧) :-

إذا علمت أن متوسط حجم الانتاج لعدد ١٠ أيام الاولى يعادل ٤٠٠ و [illegible] وأن حجم الانتاج خلال الخمسة عشر يوما التالية يعادل ١٠٥٠٠ وحده وان المعلومات المتوفرة عن حجم الإنتاج خلال الفترة المتبقية من الشهر كانت على النحو التالي :

اليوم	الاول	الثاني	الثالث	الرابع	الخامس
الإنتاج بالوحدة	٢١٠	٣٦٠	١٤٠	١٨٠	٢٠٠

المطلوب : حساب متوسط الانتاج اليومي

[illegible]

لايجاد متوسط الانتاج اليومي نعد الجدول التالي :

$\bar{س}_ر$	$ن_ر$	$ن_ر \bar{س}_ر$
$\bar{س}_1$ = ٥٠٠ وحدة	١٠	١٠ × ٥٠٠ = ٥٠٠٠
$\bar{س}_2 = \frac{مجـ\ س_ر}{ن} = \frac{١٠٥٠٠}{١٥}$ = ٧٠٠ وحدة	١٥	١٥ × ٧٠٠ = ١٠٥٠٠
$\bar{س}_3 = \frac{مجـ\ س_ر}{ن} = \frac{٢٠٠ + ١٨٠ + ١٤٠ + ٢٦٠ + ٢٢٠}{٥} = \frac{١٢٠٠}{٥}$ = ٢٤٠ وحده	٥	٥ × ٢٤٠ = ١٢٠٠
المجموع	٣٠	١٦٧٠٠

∴ متوسط الانتاج اليومى $= \frac{ن_1 \bar{س}_1 + ن_2 \bar{س}_2 + ن_3 \bar{س}_3}{ن_1 + ن_2 + ن_3}$

$= \frac{١٦٧٠٠}{٣٠}$ = ٥٥٧ وحده تقريبيا

[illegible]

١ – من النتائج التى حصلنا عليها باستخدام الطرق الثلاث السابقة يتبين لنا أن قيمة الوسط الحسابى المحسوبة واحدة أيا كانت طريقة حسابه وهى واحدة ايضا ايا كان الوسط الفرضى الذى يستخدم فى حسابه .

٢ – أن قيمة (ب) التى ورد ذكرها فى طريقة الفروق المعدلة عادة ما تختار على أنها طول الفئة فى حالة التوزيعات التكرارية المنتظمة ذات الفئات متساوية الاطوال ، أو اذا كانت أطوالها غير متساوية ولكنها تقبل القسمة على مقدار ثابت بدون باقى .

٣ – يتميز الوسط الحسابى عن غيره من المتوسطات الاخرى (الوسيط – المنوال – ....) بأنه يأخذ جميع مفردات الظاهرة فى الاعتبار عند حسابه وأنه لا يعطى الا قيمة واحدة فقط عند الحساب .

٤ – على الرغم من كافة الخصائص والمزايا التى يتمتع بها الوسط الحسابى ، الا أنه يعاب عليه عدم امكانية [illegible] فى حالة الجداول التكرارية المفتوحة [illegible] صورها الامر الذى يجعلنا نتحول عنه بايجاد بعض المتوسطات الاخرى التى لا تعتمد فى حسابها على [illegible] المفتوحة وهذا ما سنناقشه بشىء من التفصيل عند التحدث عن العلاقة بين المتوسطات المختلفة فى جزء لاحق من هذا المؤلف . كما يؤخذ عليه أيضا تأثره الشديد بالقيم المتطرفة أو الشاذة أى أنه اذا كانت احدى القيم كبيرة جدا أو صغيرة جدا بالمقارنة بباقى القيم

الاخرى فان الوسط الحسابى يتأثر بهذه القيمة مبتعدا عن القيمة الحقيقية الممثلة لباقى القيم) .

(٣-٢) الوسيط (و٢) The Median :-

ذكرنا أنه من أهم عيوب الوسط الحسابى كقيمة متوسطة هو أن حسابه يتأثر بالقيمة المتطرفة أو الشاذة اذ انه يميل الى هذه القيمة ، فالوسط الحسابى للقيمة ٢ ، ٣ ، ٥ ، ٧ ، ٨ هو ٥ بينما اذا استبدلنا المفردة الاخيرة فى المجموعة بالقيمة ٣٣ فان الوسط الحسابى يصبح ١٠ أى ضعف الوسط الحسابى الاول وذلك لمجرد تغيير قيمة واحدة بقيمة متطرفة . أضف الى ذلك أن الوسط الحسابى لا يمكن حسابه لبعض التوزيعات التكرارية (التوزيعات المفتوحة) ولهذه الاسباب مجتمعة نشأت الحاجة الى وجود مقاييس أخرى للنزعة المركزية تتلافى هذه العيوب ومن أهم هذه المقاييس ما يعرف بالوسيط .

والوسيط يعرف بأنه القيمة الوسطى أى القيمة التى تحتل المكان الاوسط بين القيم بعد ترتيبها تصاعديا أو تنازليا وبالتالى فان الوسيط عادة ما يقسم مجموعة القيم (ن) الى نصفين متساويين بحيث يكون عدد القيم الاقل من الوسيط (ن١) يساوى عدد القيم الاكبر منه (ن٢) فلا تدخل القيم المتطرفة اذا فى حسابه .

[illegible]

[illegible] ترتيبا تصاعديا (أى [illegible] قيمة ثم الاكبر منها وهكذا) أو تنازليا (أى نبدأ بأكبر قيمة ثم الاصغر منها وهكذا) ثم نبحث عن ترتيب أو رتبة الوسيط وهى الرتبة التى تقسم عدد القيم المرتبة (تصاعديا أو تنازليا) الى نصفين متساويين قبلها وبعدها .

* فاذا كان عدد القيم (ن) فرديا فان رتبة مفردة الوسيط هى $\frac{ن + ١}{٢}$ ويكون الوسيط بالتالى هو القيمة التى تحتل هذه الرتبة .

مثال (٨) :-

بفرض أنه لدينا القيم : ٤٥ ، ٥ ، ٣٩ ، ١٧ ، ٣٧

ولايجاد ترتيب الوسيط فلابد وأن نضع القيم تصاعديا مثلا أى تصبح القيم :

٥ ، ١٧ ، ٣٧ ، ٣٩ ، ٤٥

ثم نحسب ترتيب الوسيط = $\frac{ن + ١}{٢} = \frac{٥ + ١}{٢} = \frac{٦}{٢} = ٣$

أى أن الوسيط ($ر_م$) هو القيمة التى تحتل الرتبة (٣) أى [illegible] من القيم المرتبة والتى تأخذ القيمة ٣٧ (أى أن : $ر_م$ = ٣٧) وقبل هذه قيمتين هما ٥ ، ١٧ وبعدها قيمتين ايضا هما ٣٩ ، ٤٥ (لاحظ أن عدد القيم فردى) .

* اما اذا كان عدد القيم زوجيا فلا تكون هناك قيمة وسطى وحيده ولكن تكون لدينا قيمتان وسيطتان بحيث يكون عدد القيم الاقل منهما يساوى عدد القيم الاكبر منهما ويكون الوسيط هو الوسط الحسابى لهاتين القيمتين .

أى أن الخطوات فى هذه الحالة هى ترتيب القيم (تصاعديا أو تنازليا) ثم حساب ترتيب القيمتين الوسيطتين وهما : $\frac{ن}{٢}$ ، $(\frac{ن}{٢} + ١)$ ثم نحسب الوسط الحسابى للقيمتين اللتين تحملان هاتين الرتبتين فمثلا لو كانت لدينا القيم :-

٤٣ ، ٨ ، ٢٦ ، ٣٥ ، ٤٧ ، ٧٣

فبعد ترتيب هذه القيم تصاعديا نحسب رتبتى القيمتين الوسيطتين وهما

$$\frac{ن}{٢} = \frac{٦}{٢} = ٣$$

$$، \frac{ن}{٢} + ١ = ٣ + ١ = ٤$$

أى أن القيمتين الوسيطتين هما القيمتان اللتان يحتلان الترتيب الثالث والرابع وهما ٣٥ ، ٤٣

فيكون الوسيط عبارة الوسط الحسابى لهاتين القيمتين أى أن :

$$الوسيط\ (ر_٢) = \frac{٣٥ + ٤٣}{٢} = \frac{٧٨}{٢} = ٣٩$$

وقبل هذه القيمة ثلاث قيم أقل منها هى ٨ ، ٢٦ ، ٣٥ وبعدها أيضا ثلاث قيم أكبر منها هى ٤٣ ، ٤٧ ، ٧٣ .

يتضح مما سبق ان الوسيط لا يتأثر بالقيم المتطرفة اذ لا تدخل قيمتها فى حسابه ، بينما رأينا أن الوسط الحسابى يتأثر بها ولذلك فإن استخدام الوسيط فى هذه الحالة أفضل .

وهذا لا يعنى أن الوسيط يفضل دائما فأحيانا توجد بعض الصعوبات التى تعترض حسابه كما فى حالة المتغيرات المنفصلة والتى لا تأخذ قيما كسرية فاذا فرضنا أن لدينا ست أسر عدد أفرادها هى ٣ ، ٤ ، ٦ ، ٧ ، ٨ ، ١٠ واما كان عدد القيم زوجيا فان الوسيط حسب التعريف هو $\frac{٦ + ٧}{٢} = \frac{١٣}{٢} = ٦,٥$

وهذه القيمة لا وجود لها فلا معنى لأن يكون الوسيط لعدد أفراد الأسر ٦,٥ فرد وأحيانا أخرى يفقد الوسيط مدلوله اذا كان عدد المفردات صغيرا وبينها قيم كثيرة مكررة ، فعلى سبيل المثال اذا كانت لدينا درجات خمس طلاب هى ٧٠ ، ٧٠ ، ٧٠ ، ٧٢ ، ٨٠ فان الوسيط هو ٧٠ ولا توجد أى قيمة أصغر منها ويوجد ٤٠٪ من القيم أكبر منها .

ويجدر بالذكر اذا كان عدد المفردات (ن) كبيرا فسوف نتجاوز الدقة التامة فى حساب ترتيب الوسيط ونقدر الوسيط فى هذه الحالة على أنه قيمة المفردة ذات الترتيب ($\frac{ن}{٢}$) أيا كانت ن عددا فرديا أم زوجيا ، وذلك توخيا للاختصار فى الجهد الحسابى . ولعل السبب فى ذلك هو أنه فى مثل هذه الحالات يكون من اللازم علينا تبويب البيانات فى صورة جدول تكرارى ونقوم بحساب الوسيط منه كما يتضح من الفقرة التالية .

(٣-٢-٢) حساب الوسيط من البيانات المبوبة :-

اوضحنا فى فقرة سابقة أن قيمة الوسيط هى تلك القيمة التى تقسم المفردات المرتبة تصاعديا (تنازليا) الى قسمين متساويين وفى حالة التوزيعات

التكرارية لابد من اعداد جدول تكراري متجمع صاعد (هابط) وذلك لحساب قيمة الوسيط ، وسوف نعتمد على الفرض أن المفردات في الفئة الوسيطية (وهي الفئة التي تحتوي على قيمة الوسيط) تتوزع توزيعا منتظما خلال تلك الفئة . وسوف نركز في هذا المؤلف على ترتيب البيانات تصاعديا أو تكوين الجدول التكراري المتجمع الصاعد فقط عند حساب الوسيط أو الربيعين الأول والثالث .

ويمكن تعيين قيمة الوسيط باتباع الخطوات التالية :

- تكوين الجدول التكراري المتجمع الصاعد .

- نحسب ترتيب الوسيط وهي دائما = $\frac{\text{مج ك}}{٢}$

(سواء كان مجموع التكرارات فرديا أو زوجيا)

- نحدد فئة الوسيط وهي الفئة التي تقع قيمة الوسيط بين حديها الأدنى والأعلى ، حيث نبحث في عمود التكرار المتجمع الصاعد عن قيمتين متتاليتين يقع بينهما ترتيب الوسيط . هاتان القيمتان تناظران قيمتين في عمود الحدود العليا للفئات هما الحد الأدنى والأعلى للفئة الوسيطية.

- يتم حساب قيمة الوسيط بالعلاقة التالية :-

$$\text{الوسيط (ر}_٢\text{) = الحد الادنى للفئة الوسيطية} + \frac{\text{ترتيب الوسيط − التكرار الصاعد المناظر للحد الادنى لفئة الوسيط}}{\text{التكرار الصاعد المناظر للحد الاعلى لفئة الوسيط − التكرار الصاعد المناظر للحد الادنى لفئة الوسيط}} \times \text{طول الفئة الوسيطية}$$

فاذا استخدمنا الرموز التالية فان الوسيط عبارة عن :

$$\text{ر}_٢ = \text{ف د}_٢ + \frac{\frac{\text{مج ك}}{٢} - \text{ك د}_٢}{\text{ك أ}_٢ - \text{ك د}_٢} \times \text{ل}_٢$$

حيث :

$\text{ر}_٢$ = قيمة الوسيط

$\text{ف د}_٢$ = الحد الادنى للفئة الوسيطية

$\frac{\text{مج ك}}{٢}$ = ترتيب مفردة الوسيط

$\text{ك د}_٢$ = تكرار متجمع صاعد مناظر للحد الادنى للفئة الوسيطية

$\text{ك أ}_٢$ = تكرار متجمع صاعد مناظر للحد الاعلى للفئة الوسيطية

ل = طول الفئة الوسيطية .

مثال (٩) :-

احسب قيمة الوسيط لبيانات التوزيع التكرارى التالى :

فئات	١٠-	١٥-	٢٠-	٢٥-	٣٠-	٣٥-	٤٠-	٤٥-٥٠	المجموع
تكرارات	٢	٦	٩	١١	٥	٤	٣	٢	٤٠

الحل :

لإيجاد قيمة الوسيط (و) نكون جدول التوزيع التكراري المتجمع الصاعد على النحو التالي :

الجدول التكراري الأصلي		الجدول التكراري المتجمع الصاعد		ملاحظات
فئات	تكرارات	حدود عليا للفئات	تكرار متجمع صاعد	
-١٠	٢	أقل من ١٠	صفر	
-١٥	٦	أقل من ١٥	٢	
-٢٠	٦	أقل من ٢٠	٨	
-٢٥	١١	أقل من ٢٥	١٤	موقع ترتيب مفردة ر
-٣٠	٥	أقل من ٣٠	٢٥	
-٣٥	٤	أقل من ٣٥	٣٠	
-٤٠	٣	أقل من ٤٠	٣٤	
٤٥-٥٠	٣	أقل من ٤٥	٣٧	
		أقل من ٥٠	٤٠	
المجموع	٤٠	—	—	—

$$\text{ترتيب } ر_٢ = \frac{مجـ ك}{٢} = \frac{٤٠}{٢} = ٢٠$$

فى العمود التكرارى المتجمع الصاعد نجد أن ترتيب الوسيط = ٢٠ ويقع هذا الترتيب بين القيمتين ١٤ ، ٢٥ فى عمود التكرار المتجمع الصاعد وهذا يعنى أن قيمة الوسيط تقع داخل الفئة (٢٥ – ٣٠) وبالتالي فإن قيمة الوسيط يتحدد بتطبيق العلاقة الرياضية الآتية :

$$\text{ر}_{٢} = \text{ن د}_{٢} + \frac{\frac{\text{مج ك}}{٢} - \text{ك د}_{٢}}{\text{ك أ}_{٢} - \text{ك د}_{٢}} \times \text{ل}_{٢}$$

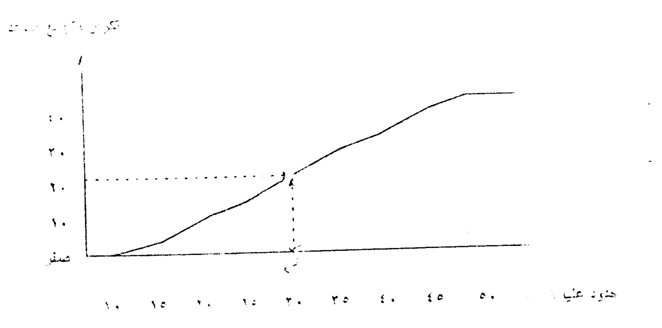
$$= ٢٥ + \frac{٢٠ - ١٤}{٢٥ - ١٤} \times ٥ = ٢٥ + \frac{٦}{١١} \times ٥ = ٢٧,٧٣ \text{ تقريبا}$$

* ايجاد الوسيط بالرسم :-

من الممكن ايجاد قيمة الوسيط بالرسم على النحو التالى :

- نرسم المنحنى التكرارى المتجمع الصاعد (أو الهابط) وذلك بأخذ الحدود العليا (أو الحدود الدنيا) للفئات على المحور الافقى والتكرار المتجمع الصاعد (أو الهابط) على المحور الرأسى كما بينا فى الباب الثانى من هذا المؤلف .
- نحدد ترتيب الوسيط وهو يساوى $\frac{\text{مج ك}}{٢}$ كما أسلفنا .
- نعين ترتيب الوسيط على المحور الراسى ، ومن هذه النقطة نرسم خطاً افقيا يقابل المنحنى الصاعد (أو الهابط) فى نقطة نسقط منها عمودا على المحور الافقى يقابله فى نقطة تكون هى قيمة الوسيط .

ففى المثال السابق يمكن حساب الوسيط بالرسم كما يلى :



من الرسم نجد أن قيمة الوسيط تساوى ١٧.٢ تقريبا ومن المؤكد أنه كلما كان الرسم دقيقا فاننا نحصل على قيمة الوسيط بدرجة دقة عالية .

<u>مميزات وعيوب الوسيط :-</u>

يتميز الوسيط عن مقياسى القيمة المتوسطة الاخرين (الوسط الحسابى - المنوال) بأنه :

- حين يزداد التواء توزيع المتغير الممثل للظاهرة موضوع الدراسة - أى بعده عن التماثل كما سنوضح فيما بعد - أى حين تتحول المشاهدات التى تسجل عن المتغير الى جانب من التوزيع دون الجانب الاخر ، فان الوسيط يفضل عن الوسط الحسابى فى هذه الحالة وذلك لأن الوسيط غالبا مالايتأثر بالقيم متطرفة متطرفة على العكس من الوسط الحسابى الذى يتأثر بوجوده فى ...

اتجاه هذه المفردات المتطرفة . وهذا هو السبب فى الاعتماد على الوسيط فى تقدير الدخل المتوسط وايضا السعر المتوسط .

- انه حين يتعذر تقدير الوسط الحسابى بدقة من توزيع تكرارى مفتوح فانه يمكن تقدير الوسيط فى مثل هذه الحالة ، هذا ما لم يكن التوزيع شديد الالتواء بحيث تقع نصف المفردات فى أى من الفئتين الاولى أو الاخيرة من التوزيع . فضلا عن أنه يمكن تقديره بالرسم .

إن الوسيط بحسب تعريفه وهو النقطة التى تقسم التوزيع الى نصفين متساويين فهو اذن مقياس ترتيبى وحين تكون القياسات التى تسجل عن ظاهرة ما من النوع الترتيبى مثل التقديرات فى أحد الاختبارات وبغض النظر عن الاسلوب الذى اتبع فى الترتيب ، فان الوسيط هو مقياس القيمة المتوسطة المناسب فى هذه الحالة وخاصة وأن الوسط الحسابى لا يحسب الا من قياسات كمية .

وخلافا لما تبين لنا من امكانية حساب الوسط الحسابى العام لعدة مجموعات فرعية وذلك باستخدام الاوساط الحسابية لهذه المجموعات ، فان هذا الامر لا يتحقق فى حالة الوسيط بمعنى أننا لا نستطيع حساب الوسيط العام لعدد من المجموعات الفرعية حتى ولو عرف الوسيط لكل مجموعة على حده وذلك بسبب عدم سهولة اخضاعه للعمليات الحسابية وبالتالى لا يمكن استخدامه لحساب متوسطات مرجحة وأخيرا يعاب على الوسيط أنه يتعذر حسابه وخاصة فى الحالات التى يقل عددها أو تتكرر بعض مفرداتها [illegible]

لا يمكن حساب الوسط ١٩ ، ١١ ، ١٢ ، ١١ ، ١١ [illegible] من الناحية العملية أن نتعامل [illegible] مع هذه الحالات [illegible] المتوسطة بالوسط الحسابي .

(٣-٣-٣) شبيهات الوسيط (الربيعين) Quartels :-

يعرف الربيع الاول أو الادنى ([illegible]) بأنه قيمة المفردة التي تقع عند نهاية الربع الاول من التوزيع ، أي التي [illegible] عندها عدد المفردات الاقل في القيمة = $\frac{١}{٤}$ عدد المفردات ، وعدد المفردات الاكبر في القيمة $\frac{٣}{٤}$ عدد المفردات . [illegible] التعريف السابق [illegible] تشابه كبير مع التعريف الخاص بالوسيط مع اختلاف الموقع على التوزيع حيث كان الوسيط هو قيمة المفردة التي تقع عند نهاية الربع الثاني من التوزيع ، أي التي يكون عندها $\frac{١}{٢}$ عدد المفردات أقل [illegible] (أكبر منها) في القيمة . [illegible] الاسلوب الذي [illegible] من قبل في تعيين قيمة الوسيط [illegible]

حساب الربيعين الاول والثالث من البيانات غير [illegible] :-

لا تختلف طريقة حساب أى من الربيعين الاول والثالث عن طريق حساب الوسيط من البيانات [illegible] الا من حيث ترتيب [illegible] ، فاذا كانت مجموعة البيانات مرتبة ترتيبا تصاعديا فان :

- ترتيب مفرده الربيع الاول = $\frac{\text{ن} + ١}{٤}$

- ترتيب مفرده الربيع الثالث = $\frac{٣ (\text{ن} + ١)}{٤}$

وذلك لعدد ن من المفردات فرديا كان أم زوجيا . ويمكن تعيين قيمتى الربيع الاول والثالث كما هو واضح من المثال التالى .

مثال (١٠) :-

حدد قيمة كل من الربيع الاول ($\text{ر}_١$) ، والوسيط ($\text{ر}_٢$) ، الربيع الثالث ($\text{ر}_٣$) للقيم التالية .

٢٠ ، ٢١ ، ١٩ ، ١٦ ، ١٤ ، ١٢ ، ٩ ، ٧ ، ١٥ ، ١١

<u>الحل</u>

أولا ترتيب القيم ترتيبا تصاعديا .

٧ ، ٩ ، ١١ ، ١٢ ، ١٤ ، ١٥ ، ١٦ ، ١٩ ، ٢٠ ، ٢١

- حساب قيمة الربيع الاول ($\text{ر}_١$) :

ترتيب مفردة الربيع الاول ($\text{ر}_١$) = $\frac{\text{ن} + ١}{٤} = \frac{١١}{٤} = ٢\frac{٣}{٤}$

والقيمة رقم ٩ (٩ ، ١١) .

∴ ر، = المفردة الثانية + $\frac{٣}{٤}$ (المفردة الثالثة − المفردة الثانية)

= ٩ + $\frac{٣}{٤}$ (١١ − ٩) = ٩ + [illegible] = ١٠ [illegible]

− حساب قيمة الربيع الثاني ([illegible])

حيث أن عدد المفردات زوجي فلا يوجد [illegible] متوسطا ترتيبيا

$\frac{ن}{٢}$ = $\frac{١٠}{٢}$ = ٥ والتالي لها = ٦

[illegible]

أي أن قيمة الوسيط = [illegible] ١٠

− حساب قيمة الربيع الثالث (ر٣)

ترتيب مفردة الربيع الثالث (ر٣) = $\frac{٣ (ن + ١)}{٤}$ = $\frac{٣ (١٠ + ١)}{٤}$

= $\frac{٣٣}{٤}$ = ٨ $\frac{١}{٤}$

اى ان مفرده $ر_٣$ تبعد عن المفرده الثامنة بمقدار $\frac{١}{٤}$ المسافة بين المفرده الثامنة والمفرده التاسعة (١٩ ، ٢٠) .

$$\therefore ر_٣ = \text{المفرده الثامنة} + \frac{١}{٤}\left(\text{المفرده التاسعة} - \text{المفرده الثامنة}\right)$$

$$= ١٩ + \frac{١}{٤}\left(٢٠ - ١٩\right) = ١٩ + \frac{١}{٤} \times ١ = ١٩\ \frac{١}{٤}$$

حساب الربيعين الاول والثالث من البيانات المبوبة .

ايضا فى هذه الحالة – تتبع نفس الخطوات التى تم اتباعها عند حساب الوسيط ، فمن جدول التوزيع التكرارى المتجمع الصاعد يلاحظ أن :

$$\text{– ترتيب مفرده الربيع الاول} = \frac{\text{مجـ ك}}{٤}$$

$$\text{– ترتيب مفرده الربيع الثالث} = \frac{\text{٣ مجـ ك}}{٤}$$

ويتم حساب قيمة الربيع الاول باستخدام العلاقة التالية :

$$ر_١ = \text{ن د}_١ + \frac{\frac{\text{مجـ ك}}{٤} - \text{ك د}_١}{\text{ك أ}_١ - \text{ك د}_١} \times ل_١$$

حيث :

$ف_{د١}$: تشير الى الحد الادنى لفئة الربيع الاول

$ك_{د١}$: تشير الى التكرار المتجمع الصاعد المناظر للحد الادنى لفئة الربيع الاول .

$ك_{أ١}$: تشير الى التكرار المتجمع الصاعد المناظر للحد الاعلى لفئة الربيع الاول .

$ل_{١}$: تشير الى طول فئة الربيع الاول .

وبالمثل فان :

$$ر_{٣} = ف_{د٣} + \frac{\frac{٣ \text{ مج ك}}{٤} - ك_{د٣}}{ك_{أ٣} - ك_{د٣}} \times ل_{٣}$$

حيث $ف_{د٣}$ ، $ك_{د٣}$ ، $ل_{أ٣}$ ، $ل_{٣}$ هى نفس التعريفات السابقة ولكن بالنسبة لفئة الربيع الثالث .

مثال (١١) :-

احسب قيمة كل من $ر_{١}$ ، $ر_{٢}$ ، $ر_{٣}$ رياضيا وبيانيا للتوزيع التكرارى التالى :

فئات	أقل من ١٠	١٠-	٢٠-	٣٠-	٤٠-	٥٠-	٦٠فاكثر
تكرارات	٤	١١	١٤	٢٧	١٨	١٧	٩

تكوين جدول التوزيع التكراري المتجمع الصاعد

الفئات	تكرارات	حدود عليا للفئات	تكرار متجمع صاعد	ملاحظات
أقل من ١٠	٤	أقل من ١٠	٤	
١٠ -	١١	أقل من ٢٠	١٥	
				ترتيب $ر_1$ = ٢٥
٢٠ -	١٤	أقل من ٣٠	٢٩	
				ترتيب $ر_2$ = ٥٠
٣٠ -	٢٧	أقل من ٤٠	٥٦	
٤٠ -	[illegible]	أقل من ٥٠	٧٤	
				ترتيب $ر_3$ = ٧٥
٥٠ -	١٧	أقل من ٦٠	٩١	
٦٠ فأكثر	٩	أقل من ∞	١٠٠	
المجموع	١٠٠			

١- حساب قيمة ($ر_1$) :

$$\text{ترتيب } ر_1 = \frac{\text{مجـ ك}}{٤} = \frac{١٠٠}{٤} = ٢٥$$

فئة $ر_1$: (٢٠ : ٣٠) طول فئة $ر_1$ $ل_1$ = ١٠

$$ر_1 = ف\,د_1 + \frac{\frac{\text{مجـ ك}}{٤} - ك\,د_1}{ك\,أ_1 - ك\,د_1} \times ل_1$$

$$\therefore ر_1 = ٢٠ + \frac{٢٥ - ١٥}{٢٩ - ١٥} \times ١٠ = ٢٧.١٤$$

٢- حساب قيمة ($ر_٢$) :

$$\text{ترتيب } ر_٢ = \frac{\text{مجـ ك}}{٢} = \frac{١٠٠}{٢} = ٥٠$$

فئة $ر_٢$: (٣٠ - ٤٠) طول فئة $ر_٢$ = $ل_٢$ = ١٠

$$ر_٢ = \text{ف } د_٢ + \frac{\frac{\text{مجـ ك}}{٢} - \text{ك } د_٢}{\text{ك } أ_٢ - \text{ك } د_٢} \times ل_٢$$

$$\therefore ر_٢ = ٣٠ + \frac{٥٠ - ٢٩}{٥٦ - ٢٩} \times ١٠ = ٣٧,٧٨$$

٣- حساب قيمة ($ر_٣$) :

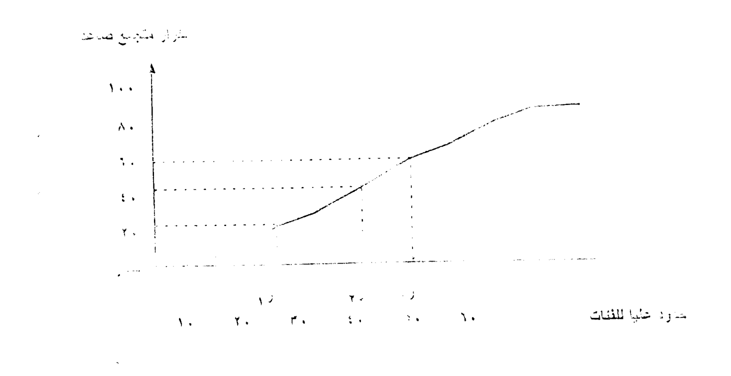
$$\text{ترتيب } ر_٣ = \frac{\text{٣مجـ ك}}{٤} = \frac{٣ \times ١٠٠}{٤} = ٧٥$$

فئة $ر_٣$: (٥٠ - ٦٠) طول فئة $ر_٣$ = $ل_٣$ = ١٠

$$ر_٣ = \text{ف } د_٣ + \frac{\frac{\text{٣مجـ ك}}{٤} - \text{ك } د_٣}{\text{ك } أ_٣ - \text{ك } د_٣} \times ل_٣$$

$$\therefore ر_٣ = ٥٠ + \frac{٧٥ - ٧٤}{٩١ - ٧٤} \times ١٠ = ٥٠,٥٩$$

ويمكن تحديد المقـاييس الثلاثـة السـابقة ($ر_١$ ، $ر_٢$ ، $ر_٣$) بالرسم ، وذلك برسم المنحنى التكرارى المتجمع الصاعد على النحو التالى



من الرسم نجد أن :

ر١ ، ر٢ ، ر٣ هى نفس القيم المحسوبة تقريبا .

<u>[illegible]</u>

<u>(أ) حساب المنوال فى حالة البيانات الغير مبوبة</u>

يعرف المنوال بمجموعة من القيم بانه الاكثر شيوعا – أى القيمة التى تكررت أكثر من غيرها .

<u>فمثلا :</u> اذا كان لدينا القيم الاتية : ٤ ، ٥ ، ٦ ، ٧ ، ٥ ، ٨ ، ٥ ، ٦ فان المنوال لهذه المجموعة هو القيمة (٥) لأنها تكرر أكثر من غيرها .

الا أنه يلاحظ من التعريف السابق للمنوال انه لا يمكن حسابه فى الحالات التى لا يحدث فيها تكرار لقيمة معينة عن باقى قيم الظاهرة ، كذلك يلاحظ حصولنا فى بعض الحالات على اكثر من قيمة للمنوال وذلك اذا ما تكررت هذه القيم بعدد متساوى .

وعموما ، ينبغى ملاحظة أن دور المنوال فى علم الاحصاء لا يرقى الى دور الوسط الحسابى أو الوسيط كمقياس للنزعة المركزية ، اذ يهتم بالمنوال عادة أصحاب الاعمال الصناعية أو التجارية ، فالقيمة الاكثر تكرار لها مغزى خاص بالنسبة لهم فالمنتج الاكثر رواجا فى صناعة ما يجذب اهتمام اصحاب هذه الصناعة بزيادة المنتج منها وعمل الدعاية اللازمة لها ، كما أن المهتمون بالعلوم السلوكية عادة ما يعتمدون على المنوال فى دراسة القيمة المتوسطة لتلك الظواهر خاصة وأن المنوال قابل للحساب فى جميع انواع البيانات .

(٣-٣-٢) حساب المنوال فى حالة البيانات المبوبة .

يتطلب حساب المنوال فى حالة البيانات المبوبة توفيق أحسن منحنى يمثل التوزيع الذى لدينا ومنه نوجد القيمة التى تناظر قمته فتكون هى قيمة المنوال . وتوفيق أحسن منحنى يتطلب ضرورة ايجاد معادلة هذا المنحنى حتى نستطيع الحصول على قيمة المنوال بدقة وهذا ليس مجالنا فى هذه الدراسة ، الا انه يمكن ايجاد قيمة المنوال سواء بالحساب أو بالرسم بطرق تقريبية منها :

(٤) طريقة الرافعة :

وتقوم هذه الطريقة على أساس أن المنوال طالما هو القيمة الأكثر تكرارات فهو يقع فى الفئة ذات التكرار الاكبر ، وهذه الفئة تعرف باسم " الفئة المنوالية " ولتحديد موقع المنوال داخل هذه الفئة نفترض مبدئيا أن المنوال يقع فى مركز هذه الفئة ، ولكن هذا يعتبر تقريبيا ولا يكون صحيحا الا اذا كان التوزيع متماثلا تماما . أما اذا لم يكن التوزيع متماثلا وهو الغالب فان المنوال ينحرف عن مركز الفئة نحو بدايتها أو نهايتها قليلا أو كثيرا حسب شدة الاختلاف بين قيمتى التكرارين للفئتين السابقة واللاحقة للفئة المنوالية وبذلك يمكن ايجاد قيمة المنوال من العلاقة الاتية :

قيمة المنوال = بداية الفئة المنوالية (المناظرة لاكبر تكرار) + س

حيث يمكننا حساب قيمة المجهول س من العلاقة الاتية :

التكرار السابق للفئة المنوالية × س = التكرار اللاحق للفئة المنوالية × (طول الفئة المنوالية − س)

ويؤخذ على هذه الطريقة عدم دقتها لانها تسقط أكبر تكرار فى التوزيع وهو تكرار الفئة المنوالية .

(٢) طريقة الفروق (طريقة بيرسون) :

وتقوم هذه الطريقة على أساس تلافى العيب الموجود فى طريقة الرافعة وهو اهمال أكبر تكرار بالتوزيع (تكرار الفئة المنوالية) عند حساب قيمة المنوال . فهذه الطريقة تعتمد على تكرارات الفئة المنوالية والفئتين المحيطتين بها وذلك بأخذ الفروق بين تكرار الفئة المنوالية وتكرارات هاتين الفئتين وبذلك يمكن ايجاد قيمة المنوال من العلاقة الاتية :

قيمة المنوال = بداية الفئة المنوالية (المناظرة لاكبر تكرار) + س

حيث يمكننا حساب قيمة المجهول س من العلاقة الأتية :

$$\frac{\text{س}}{\text{طول الفئة المنوالية} - \text{س}} = \frac{\text{تكرار الفئة المنوالية} - \text{التكرار السابق له}}{\text{تكرار الفئة المنوالية} - \text{التكرار اللاحق له}}$$

ولاشك أن هذه الطريقة ادق من طريقة الرافعة لأنها تأخذ فى اعتبارها تكرار الفئة المنوالية ذاتها بالاضافة الى تكرارى الفئتين المحيطتين بها .

مثال (١٢) :-

أخذت عينة مكونة من ٢٠٠ من المشتغلين باحدى الوزارات فكان توزيعهم العمرى كما يلى :

[illegible] بالسنوات	٢٠ -	٢٥ -	٣٠ -	٣٥ -	٤٠ -	٤٥ -	٥٠ -	٥٥ - ٦٠	المجموع
عدد المشتغلين	١٠	١٧	٢٤	٤٣	٣٤	٣٠	٢٣	١٦	٢٠٠

المطلوب :

ايجاد قيمة المنوال بطريقتى الرافعة والفروق

الحل

أولا حساب المنوال بطريقة الرافعة :

يلاحظ بالنسبة للتوزيع التكرارى السابق أن :

- التوزيع منتظم (أطوال فئاته متساوية) وبالتالى يمكن استخدام التكرارات الاصلية فى ايجاد قيمة المنوال .

- الفئة المنوالية (التى تناظر أكبر تكرار) هى ٣٥ - ٤٠

- طول الفئة المنوالية = ٤٠ - ٣٥ = ٥

∴ قيمة المنوال = بداية الفئة المنوالية + س

= ٣٥ + س

حيث قيمة س تحسب من العلاقة :

التكرار السابق للفئة المنوالية × س = التكرار اللاحق للفئة المنوالية × (طول الفئة المنوالية - س)

∴ ٢٤ × س = ٣٤ (٥ - س)

٢٤س = ١٧٠ - ٣٤س

$\therefore$ ٥٨س = ١٧٠

$\therefore$ $س = \frac{١٧٠}{٥٨} = ٢,٩٣$

$\therefore$ م = ٣٥ + ٢,٩٣ = ٣٧,٩٣

ثانيا حساب قيمة المنوال بطريقة الفروق :

من المعلومات المتوفرة عن التوزيع التكرارى السابق

$\therefore$ قيمة المنوال = بداية الفئة المنوالية + س

= ٣٥ + س

حيث قيمة س تحسب من العلاقة :

$$\frac{س}{\text{طول الفئة المنوالية} - س} = \frac{\text{تكرار الفئة المنوالية} - \text{التكرار السابق لها}}{\text{تكرار الفئة المنوالية} - \text{التكرار اللاحق لها}}$$

$\therefore$ $\frac{س}{٥ - س} = \frac{٤٣ - ٢٤}{٤٣ - ٣٤}$

$\therefore$ $\frac{س}{٥ - س} = \frac{١٩}{٩}$

$\therefore$ ٩ س = ٩٥ - ١٩س

$\therefore$ ٢٨ س = ٩٥

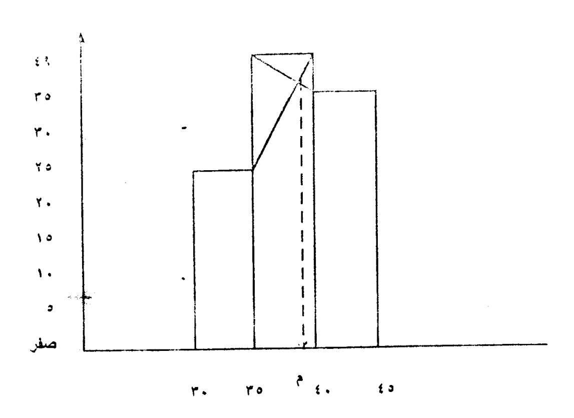
$\therefore$ س = ٣,٣٩

م = ٣٥ + ٣,٣٩ = ٣٨,٣٩

وليس غريبا أن نحصل على نتيجتين مختلفتين باستخدام الطريقتين السابقتين وهذا هو الواقع الطبيعى : اذ أن هاتين الطريقتين تقومان على أساسين مختلفتين والواقع أن كليهما تقريبى وان كانت احداهما وهى طريقة الفروق أقرب الى الدقة من الاخرى .

<u>* حساب المنوال بالرسم (طريقة الفروق) :-</u>

يتم حساب قيمة المنوال بالرسم من المدرج التكرارى ، وإن كان يكتفى برسم المستطيلات التى تمثل الفئة المنوالية والفئة السابقة واللاحقة لها ، ثم نصل الرأسى الايمن العلوى من مستطيل الفئة المنوالية بالرأسى الايمن العلوى للمستطيل السابق له (الذى يمثل الفئة السابقة للفئة المنوالية) وكذلك نصل الرأسى الايسر العلوى لمستطيل الفئة المنوالية بالرأسى الايسر العلوى للمستطيل اللاحق له (الذى يمثل الفئة اللاحقة للفئة المنوالية) فيتقاطعان فى نقطة نسقط من عندها عمودا على المحور الافقى يقابله فى نقطة تكون هى قيمة المنوال ، كما يتضح من الشكل الذى يوضح كيفية ايجاد قيمة المنوال بالرسم من المدرج التكرارى للمثال السابق



من الرسم يتضح أن قيمة المنوال (م) = ٣٨,٣ تقريبا

ويجب ملاحظة أنه اذا كانت فئات التوزيع غير متساوية (التوزيع غير منتظم) فلابد من تعديل التكرارت قبل حساب المنوال سواء بالحساب بالطريقتين (الرافعة أو الفروق) أو بالرسم ، وهو مالم يحدث فى الوسط الحسابى أو الوسيط ، وذلك لأن قيمة المنوال تحسب باستخدام المدرج التكرارى الذى يتطلب ضرورة تعديل التكرارت قبل رسمه .

فى حالة التوزيعات الغير المنتظمة حيث : (التكرار المعدل = $\frac{\text{التكرار الاصلى}}{\text{طول الفئة}}$)

وبعد تعديل التكرارات نستخدم هذه التكرارات المعدلة فى حساب قيمة المنوال سواء بالحساب أو بالرسم كالمعتاد .

ومن ناحية أخرى فان المنوال يختلف عن الوسط الحسابى حيث يمكن حسابه من الجداول التكرارية المفتوحة شأنه فى ذلك شأن الوسيط والربعين الاول والثالث حيث أنه يعتمد فى حسابه على التكرارت فقط وليس على مراكز الفئات كما هو الحال فى ايجاد الوسط الحسابى .

العلاقة بين الوسط الحسابى والوسيط والمنوال وشكل التوزيع

يعرف التوزيع المتماثل بأنه ذلك التوزيع الوحيد القمة (أى ذو قيمة منوالية وحيدة) الذى يقسمه محور تماثله (أى العمود الساقط من قمة التوزيع المحور الافقى) الى قسمين متماثلين وكأن أحد القسمين صورة معكوسة للقسم الاخر .

أما الالتواء فيقيس مقدار البعد عن هذا الشكل (المتماثل) . ويجب الاشارة أن ما يعنينا فى هذا الجزء فقط هو الربط بين شكل التوزيع ومقاييس القيمة المتوسطة له وتتلخص هذه العلاقة فى أن المقاييس الثلاث تتطابق حين يكون التوزيع متماثل أى أن :

($\bar{س}$ = ر$_2$ = م) .

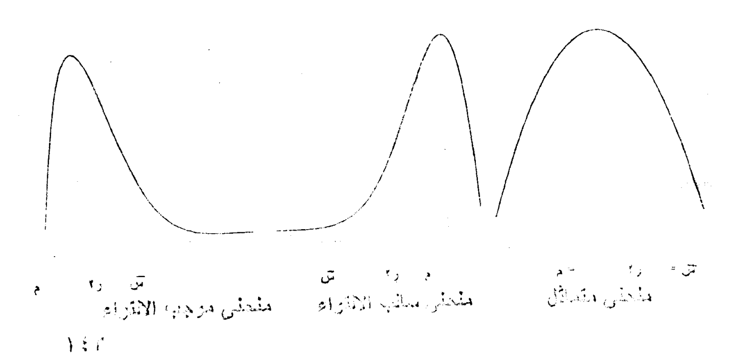
أما اذا كان المنحنى بسيط الالتواء أو قريب من التماثل فان الوسيط يقع دائما بين الوسط الحسابى والمنوال . ونحن بهذا الصدد فانه يجب أن نشير الى أن

[illegible] = ٣ (س̄ - ر)

يلاحظ أنه يمكن استخدام العلاقة السابقة فى ايجاد مقياس من المقاييس الثلاثة بمعلومية المقياسان الآخران . كما أنه يمكن الاستفادة من هذه العلاقة بصفة خاصة فى تقدير قيمة الوسط الحسابى فى حالة التوزيعات التكرارية المفتوحة . حيث يتم حساب كل من الوسيط والمنوال بهذا التوزيع التكرارى المفتوح مباشرة (لأن الوسيط والمنوال لا يعتمدان فى حسابهما على مراكز الفئات) ثم بالتطبيق فى العلاقة السابقة بين المتوسطات الثلاثة يمكننا تقدير الوسط الحسابى .

والاشكال الثلاثة التالية (أ ، ب ، جـ) توضح كيفية استخدام العلاقة السابقة فى التعرف على شكل المنحنى :

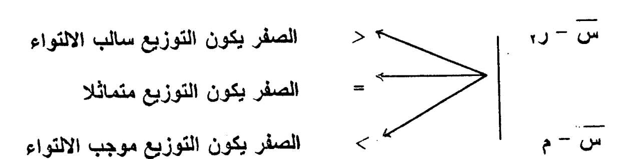
شكل أ　　　شكل ب　　　شكل ج



س̄ = ر = م　　　م　ر　س̄　　　س̄　ر　م

منحنى متماثل　　　منحنى سالب الالتواء　　　منحنى موجب الالتواء

مما سبق نستطيع أن نستنتج العلاقة بين مقاييس القيمة المتوسطة الثلاث وشكل التوزيع على النحو التالى :



وهناك علاقة أخرى تربط بين الوسيط والربعين وكيفية الاستفادة منها فى التعرف على شكل التوزيع وهى كالاتى :

$(ر_٣ - ر_٢) - (ر_٢ - ر_١)$

< الصفر يكون التوزيع سالب الالتواء

= الصفر يكون التوزيع متماثلا

> الصفر يكون التوزيع موجب الالتواء

وسوف نكتفى بهذا القدر فى مناقشة العلاقة بين المتوسطات الثلاث فى الوقت الحاضر ولنا عودة فى جزء لاحق من هذا المؤلف .

(٣-٤) الوسط الهندسى *GEOMETRIC MEAN* :-

(٣-٤-١) حسابه فى حالة البيانات غير المبوبة .

الوسط الهندسى لمجموعة من القيم عددها (ن) ، هو الجذر النونى لحاصل ضرب هذه القيم ، فاذا فرضنا أن $س_١$ ، $س_٢$ ، $س_٣$ ، ..... ، $س_ن$ هى قيم

فالاشارة هنا ، فان الوسط الهندسى - عادة ما يرمز له بالرمز (هـ) فيه الآتى

يكون

$$هـ = \sqrt[ن]{س_{1} \times س_{2} \times س_{3} \times ..... \times س_{ن}}$$

ولتسهيل العمليات الحسابية الخاصة بايجاد قيمة الوسط الهندسى ، نـأخذ لوغاريتم الطرفين ، أى أن :

$$لو\ هـ = \frac{1}{ن} (لو\ س_{1} + لو\ س_{2} + لو\ س_{3} + \cdots + لو\ س_{ن})$$

$$= \frac{مجـ\ لو\ س}{ن}$$

أى أن لوغاريتم الوسط الهندسى لمجموعة من القيم يساوى الوسط الحسابى للوغاريتمات هذه القيم . وبالكشف فى جدول الاعداد المقابلة للوغاريتمات نحصل على قيمة الوسط الهندسى (هـ) .

مثال (١٣) :-

اذا كانت أطوال خمسة من الطلاب (بالسنتيمتر) فى احدى المدارس الثانوية هى ١٥٥ ، ١٦٠ ، ١٦٣ ، ١٦٤ ، ١٦٨ فان الوسط الهندسى لهذه الاطوال يمكن حسابه كما يلى :

$$هـ = \sqrt[5]{155 \times 160 \times 163 \times 164 \times 168}$$

$$لو\ هـ = \frac{1}{5}(لو\ 155 + لو\ 160 + لو\ 163 + لو\ 164 + لو\ 168)$$

$$= \frac{1}{5}(2.1903 + 2.2041 + 2.2122 + 2.2148 + 2.2253)$$

$$= \frac{1}{5}(11.0467) = 2.2093$$

وبالكشف فى جدول الاعداد المقابلة نجد أن :

الوسط الهندسى (هـ) = ١٦١,٩ سم

(٣-٤-٣) حسابه فى حالة البيانات المبوبة (التوزيعات التكرارية)

يلاحظ عند ايجاد الوسط الهندسى للبيانات المبوبة أننا نستخدم الصيغة السابقة وفى حالة البيانات الغير مبوبة ، ولكن يأخذ الصورة التالية :

$$هـ = \sqrt[مجـ\ ك]{س_1^{ك_1} \times س_2^{ك_2} \times \ldots\ldots \times س_ن^{ك_ن}}$$

حيث $س_ر$: تمثل مركز الفئة رقم ر

$ك_ر$: تمثل التكرار المناظر للفئة رقم ر

ولتسهيل العمليات الحسابية نحول الصيغة السابقة الى الصيغة التالية :

$$\text{لو هـ} = \frac{1}{\text{مجـ ك}} \left(\text{ك}_1 \text{ لو س}_1 + \text{ك}_2 \text{ لو س}_2 + \ldots + \text{ك}_{\text{ن}} \text{ لو س}_{\text{ن}} \right)$$

$$= \frac{\text{مجـ ك}_{\text{ر}} \text{ لو س}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}}$$

وبالكشف فى جاول الاعداد المقابلة للوغاريتمات نحصل على قيمة الوسط الهندسى (هـ) .

وعند حسابه نستخدم الخطوات الاتية :

١. نعين مراكز الفئات $\text{س}_{\text{ر}}$ كالمعتاد .

٢. نوجد لوغاريتم مراكز الفئات .

٣. نضرب لوغاريتم مراكز الفئات فى التكرارات المقابلة لها ثم نوجد مجموع حواصل الضرب (مجـ ك لو س) .

٤. نقسم هذا المجموع على مجموع التكرارات (مجـ ك) فنحصل على لوغاريتم الوسط الهندسى (لو هـ) .

٥. باستخدام جداول الاعداد المقابلة للوغاريتمات نحصل على قيمة الوسط الهندسى المطلوب (هـ) .

مثال [illegible]

الجدول التالى يبين التوزيع التكرارى لدرجات ٦٢ طالب فى أحد امتحانات مادة الاحصاء فى الفرقة الثالثة بكلية التجارة :

فئت الدرجات	١٢-	١٦-	٢٠-	٢٤-	٢٨-	٣٢-	٣٦-٤٠
عدد الطلاب	٥	٨	١٢	١٨	١٠	٧	٤

أوجد الوسط الهندسى لدرجات الطلبة فى هذا الامتحان .

الحل :

لايجاد الوسط الهندسى لدرجات الطلاب ننشأ الجدول التالى :

ذ	عدد الطلبة (التكرار = ك)	مراكز الفئات س	لوغاريتم مراكز الفئات (لو س)	ك × لو س
١٢-	٥	١٤	١,١٤٦١	٥,٧٣٠٥
١٦-	٨	١٨	١,٢٥٥٣	١٠,٠٤٢٤
٢٠-	١٢	٢٢	١,٣٤٢٤	١٦,١٠٨٨
٢٤-	١٨	٢٦	١,٤١٥٠	٢٥,٤٧٠٠
٢٨-	١٠	٣٠	١,٤٧٧١	١٤,٧٧١٠
٣٢-	٧	٣٤	١,٥٣١٥	١٠,٧٢٠٥
٣٦-٤٠	٤	٣٨	١,٥٧٩٨	٦,٣١٩٢
المجموع	٦٤	———	———	٨٩,١٦٢٤

$$= \frac{\sum ك \, ر}{\sum ك} = \frac{٨٩,١٦٢٤}{٦٤} = ١,٣٩٣٢$$

وبالكشف فى جدول الاعداد المقابلة للوغاريتمات نجد أن :

الوسط الهندسى (هـ) = ٢٤,٧٣ درجة .

ومما ينبغى ملاحظته فان الوسط الهندسى للتوزيعات التكرارية غير المنتظمة لا يتطلب تعديل التكرارات شأنه فى ذلك شأن الوسط الحسابى لهذه التوزيعات كما انه لا يمكن حسابه من جداول تكرارية مفتوحة وذلك للاعتماد على مراكز الفئات عند حسابه .

(٣-٥) الوسط التوافقى *HARMONIC MEAN* :-

(٣-٥-١) حسابه فى حالة البيانات الغير مبوبة :

الوسط التوافقى لمجموعة من القيم عددها (ن) هو مقلوب الوسط الحسابى لمقلوبات هذه القيم .

فاذا فرضنا أن $س_١$ ، $س_٢$ ، ٠٠٠٠٠٠ ، $س_ن$ هى قيم ظاهرة ما : فان الوسط التوافقى والذى عادة ما يرمز له بالرمز (ق) لهذه القيم يكون :

$$ق = \frac{ن}{\sum \left(\frac{١}{س} \right)}$$

ولحساب الوسط التوافقى نتبع الخطوات الآتية :

١.نوجد مقلوبات القيم ($\frac{\text{١}}{\text{س}}$) ثم نوجد حاصل جمعها (مجـ $\frac{\text{١}}{\text{س}}$)

٢.نقسم عدد القيم (ن) على المجموع السابق فنحصل على الوسط التوافقى (ق)

ففى المثال السابق مباشرة للبيانات غير المبوبة يلاحظ أن :

$$\text{ق} = \frac{\text{ن}}{\text{مجـ } \left(\frac{\text{١}}{\text{س}}\right)}$$

$$= \frac{\text{٥}}{\frac{\text{١}}{\text{١٥٥}} + \frac{\text{١}}{\text{١٦٠}} + \frac{\text{١}}{\text{١٦٣}} + \frac{\text{١}}{\text{١٦٤}} + \frac{\text{١}}{\text{١٦٨}}}$$

$$= \frac{\text{٥}}{\text{٠,٠٠٦٤٥} + \text{٠,٠٠٦٢٥} + \text{٠,٠٠٦١٤} + \text{٠,٠٠٦١٠} + \text{٠,٠٠٥٩٥}}$$

$$= \frac{\text{٥}}{\text{٠,٠٣٠٨٩}} = \text{١٦١,٨ سم}$$

(٣-٥-٣) حسابه فى حالة البيانات المبوبة (التوزيعات التكرارية)

يستخدم القانون السابق فى حالة البيانات الغير مبوبة مع اجراء تعديل على النحو التالى :

$$\bar{\text{ق}} = \frac{\text{مج}\ \text{ك}_{\text{ر}}}{\text{مج} \left(\frac{\text{ك}_{\text{ر}}}{\text{س}_{\text{ر}}} \right)}$$

حيث س : هي مراكز الفئات

ك : هي التكرارات المناظرة لكل فئة

وعند حسابه نتبع الخطوات التالية :

١ – نعين مراكز الفئات لكل فئة كالمعتاد .

٢ – نقسم تكرار كل فئة على المركز المناظر لها فنحصل على $\left(\frac{\text{ك}_{\text{ر}}}{\text{س}_{\text{ر}}} \right)$

٣ – نوجد مجموع هذه القيم فنحصل على مج $\left(\frac{\text{ك}_{\text{ر}}}{\text{س}_{\text{ر}}} \right)$.

٤ – نقسم مجموع التكرارات (مج $\text{ك}_{\text{ر}}$) على المجموع السابق فنحصل على الوسط التوافقي ($\bar{\text{ق}}$) .

في المثال السابق مباشرة (بيانات مبوبة) احسب الوسط التوافقي لدرجات الطلاب .

الحل

لايجاد الوسط التوافقي لدرجات الطلاب ننشأ الجدول التالي :

التكرار / مراكز الفئات $(\frac{ك_ر}{س_ر})$	مراكز الفئات $س_ر$	عدد الطلبة (التكرار = ك)	فئات الدرجات
٠,٣٦	١٤	٥	١٢-
٠,٤٤	١٨	٨	١٦-
٠,٥٥	٢٢	١٢	٢٠-
٠,٦٩	٢٦	١٨	٢٤-
٠,٣٣	٣٠	١٠	٢٨-
٠,٢١	٣٤	٧	٣٢-
٠,١١	٣٨	٤	٣٦-٤٠
٢,٦٩			المجموع

$$ق = \frac{مجـ ك_ر}{مجـ (\frac{ك_ر}{س_ر})}$$

$$= \frac{٦٤}{٢,٦٩} = ٢٣,٧٩ \text{ درجة}$$

يلاحظ أيضا أن مثل الوسط الحسابى والهندسى لا يستدعى حسابه تعديل التكرارات فى حالة التوزيعات غير المنتظمة كما انه لا يمكن ايجاده من التوزيع التكرارى المفتوح وذلك للاعتماد على مراكز الفئات عند حسابه .

مثال عام :-

الجدول الاتى يوضح توزيع عينة من الاسر فى احدى المدن حسب فئات الانفاق الشهرى بالجنيه .

فئات الانفاق الشهرى	٥-	٢٥-	٤٥-	٦٥-	٨٥-	١٠٥-	١٢٥-	١٤٥-١٦٥
عدد الاسر	٤	٦	١٥	٢٢	١٣	٧	٨	٠

المطلوب :

١- حساب قيمة الوسط الحسابى لانفاق هذه الاسر .

٢- حساب قيمة الوسيط والربيع الاول والثالث رياضيا مع شكل التوزيع ثم تحديدهم بيانيا .

٣- حساب قيمة المنوال .

(أ) بالحساب بطريقتى الرافعة والفروق .

(ب) بالرسم من المدرج التكرارى .

٤- حساب قيمة الوسط الهندسى .

٥- حساب قيمة الوسط التوافقى .

الحل :

١- حساب قيمة الوسط الحسابى (سَ)

لايجاد قيمة الوسط الحسابى ننشأ الجدول التالى :

أ = ٧٥   ب = ٢٠

فئات الانفاق	عدد الاسر (التكرار = ك)	مراكز الفئات ($س_ر$)	ح = ($س_ر$ − ٧٥)	$حَ_ر = \frac{ح}{٢٠}$	$حَ_ر ك_ر$
٥-	٤	١٥	٦٠-	٣-	١٢-
٢٥-	٦	٣٥	٤٠-	٢-	١٢-
٤٥-	١٥	٥٥	٢٠-	١-	١٥-
٦٥-	٢٢	٧٥	صفر	صفر	صفر
٨٥-	١٣	٩٥	٢٠	١	١٣
١٠٥-	٧	١١٥	٤٠	٢	١٤
١٢٥-	٥	١٣٥	٦٠	٣	١٥
١٤٥-١٦٥	٣	١٥٥	٨٠	٤	١٢
المجموع	٧٥	-	-		٣٩- ٥٤+ ——— ١٥+

∴ الوسط الحسابي ($\overline{س}$) = أ + $\frac{مج\ \bar{ح}\ ك}{مج\ ك}$ × ب

= ٧٥ + $\frac{١٥}{٧٥}$ × ٢٠ = ٧٩ جنيها

يلاحظ أننا في حل هذا التمرين استخدمنا طريقة (الفروق المعدلة) لأن الفئات هنا متساوية في الطول . كذلك اتخذنا الوسط الفرضي أمام أكبر تكرار وهو (٢٢) تسهيلا للعمليات الحسابية .

٢ – حساب قيمة الوسيط (ر٢):

(أولا) حساب قيمة الوسيط رياضيا (من الجدول التكراري المتجمع الصاعد)

لإيجاد قيمة الوسيط (ر٢) رياضيا لابد أولا من انشاء الجدول التكراري المتجمع الصاعد كالمعتاد على النحو التالي :

الجدول التكراري الأصلي		الجدول التكراري المتجمع الصاعد		ملاحظات
فئات	تكرارات	حدود عليا للفئات	تكرار متجمع صاعد	
٥-	٤	أقل من ٢٥	٤	
٢٥-	٦	أقل من ٤٥	١٠	
٤٥-	١٥	أقل من ٦٥	٢٥	موقع ترتيب ر١
٦٥-	٢٢	أقل من ٨٥	٤٧	موقع ترتيب ر٢
٨٥-	١٣	أقل من ١٠٥	٦٠	موقع ترتيب ر٣
١٠٥-	٧	أقل من ١٢٥	٦٧	
١٢٥-	٥	أقل من ١٤٥	٧٢	
١٤٥-١٦٥	٣	أقل من ١٦٥	٧٥	
المجموع	٧٥			

أ- ايجاد قيمة الوسيط ($ر_٢$)

$$\text{ترتيب الوسيط} = \frac{\text{مجـ ك}}{٢} = \frac{٧٥}{٢} = ٣٧,٥$$

∴ الفئة الوسيطية هى (٦٥-٨٥) ، طول الفئة الوسيطية ($ل_٢$) = ٢٠

$$∴ ر_٢ = \text{ف د}_٢ + \frac{\frac{\text{مجـ ك}}{٢} - \text{ك د}_٢}{\text{ك أ}_٢ - \text{ك د}_٢} \times ل_٢$$

$$= ٦٥ + \frac{٣٧,٥ - ٢٥}{٤٧ - ٢٥} \times ٢٠ = ٧٦,٣٦ \text{ جنيها}$$

ب - ايجاد قيمة الربيع الاول ($ر_١$)

$$\text{ترتيب الربيع الاول} = \frac{\text{مجـ ك}}{٤} = \frac{٧٥}{٤} = ١٨,٧٥$$

∴ فئة الربيع الاول هى (٤٥- ٦٥) ، طول فئة الربيع الاول = ٢٠ .

$$∴ ر_١ = \text{ف د}_١ + \frac{\frac{\text{مجـ ك}}{٤} - \text{ك د}_١}{\text{ك أ}_١ - \text{ك د}_١} \times ل_١$$

$$= ٤٥ + \frac{١٨,٧٥ - ١٠}{٢٥ - ١٠} \times ٢٠ = ٥٦,٦٧ \text{ جنيها}$$

جـ - ايجاد قيمة الربيع الثالث ($ر_٣$)

$$\text{ترتيب الربيع الثالث} = \frac{٣\text{مجـ ك}}{٤} = \frac{٣\times ٧٥}{٤} = ٥٦,٢٥$$

∴ فئة الربيع الثالث هى (٨٥ – ١٠٥) ، طول فئة الربيع الثالث = ٢٠

$$\therefore ر_٣ = ف\,د_٣ + \frac{\frac{٣مجـ\,ك}{٤} - ك\,د_٣}{ك\,أ_٣ - ك\,د_٣} \times ل_٣$$

$$= ٨٥ + \frac{٥٦,٢٥ - ٤٧}{٦٠ - ٤٧} \times ٢٠ = ٩٩,٢٣ \text{ جنيها}$$

٤ – تعيين شكل التوزيع الانفاقى الشهرى بناء على المعلومات المتوفرة (ر١ ، ر٢ ، ر٣)

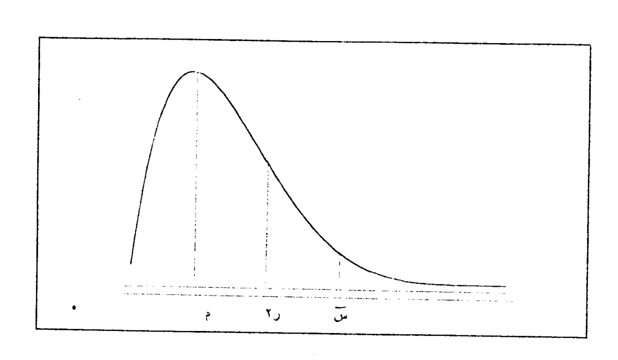
يمكن تعيين شكل الالتواء لتوزيع الانفاق الشهرى باستخدام العلاقة التالية:

(ر٣ – ر٢) – (ر٢ – ر١)	< صفر	التوزيع سالب الالتواء
	= صفر	التوزيع متماثل
	> صفر	التوزيع موجب الالتواء

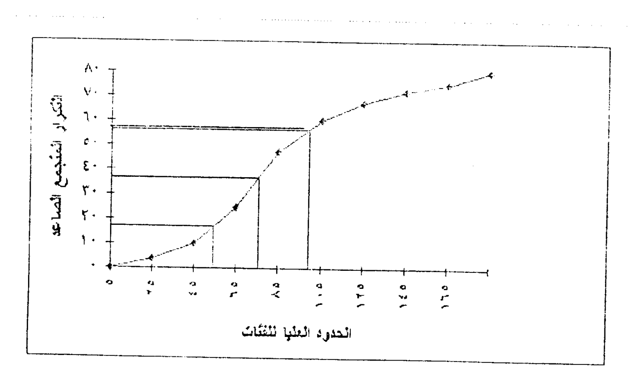
وحيث أن $ر_١ = ٥٦,٦٧$ ، $ر_٢ = ٧٦,٣٦$ ، $ر_٣ = ٩٩,٢٣$

∴ (٩٩,٢٣ – ٧٦,٣٦) – (٧٦,٣٦ – ٥٦,٦٧) = + ٣,١٧

وحيث أن ناتج التعويض كمية موجبة (أكبر من الصفر) ، فإن منحنى **توزيع الانفاق الشهرى للاسر هو منحنى موجب الالتواء ويأخذ الشكل التالى :**



(ثانيا) تعيين ر١ ، ر٢ ، ر٣ من المنحنى المتجمع الصاعد



٣- حساب قيمة المنوال (م)

(أ) بالحساب باستخدام طريقة الرافعة

م = بداية الفئة المنوالية + س

حيث س يتم حسابها كما يلى :

التكرار السابق للفئة المنواليـة × س = التكرار اللاحـق للفئة المنواليـة (طول الفئة المنوالية – س)

∴ ١٥ × س = ١٣ (٢٠ – س)

١٥ س = ٢٦٠ – ١٣ س

٢٨ س = ٢٦٠

∴ س = $\frac{٢٦٠}{٢٨}$ = ٩,٣

∴ قيمة المنوال (م) = ٦٥ + ٩,٣ = ٧٤,٣

(ب) بالحساب باستخدام طريقة الفروق (بيرسون) :

قيمة المنوال = بداية الفئة المنوالية + س

= ٦٥ + س

حيث س يتم حسابها كما يلى :

$\frac{\text{س}}{\text{طول الفئة المنوالية – س}}$ = $\frac{\text{تكرار الفئة المنوالية – التكرار السابق لها}}{\text{تكرار الفئة المنوالية – التكرار اللاحق لها}}$

$$\therefore \frac{\text{س}}{\text{٢٠} - \text{س}} = \frac{\text{٢٢} - \text{١٥}}{\text{٢٢} - \text{١٣}}$$

$$\frac{\text{س}}{\text{٢٠} - \text{س}} = \frac{\text{٧}}{\text{٩}}$$

$$\therefore \text{٩س} = \text{٧} (\text{٢٠} - \text{س})$$

$$\text{٩س} = \text{١٤٠} - \text{٧س}$$

$$\text{١٦س} = \text{١٤٠}$$

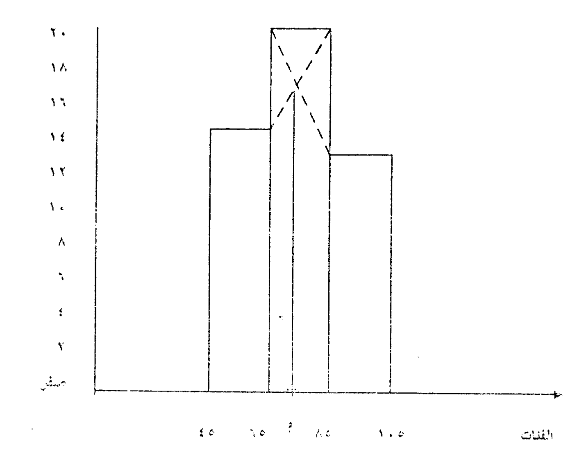
$$\therefore \text{س} = \frac{\text{١٤٠}}{\text{١٦}} = \text{٨,٧٥}$$

$$\therefore \text{م} = \text{٦٥} + \text{٨,٧٥} = \text{٧٣,٧٥} \text{ جنيها}$$

يلاحظ أننا حصلنا على قيمتين مختلفتين للمنوال وهذا أمر طبيعي نظرا لاختلاف الأساس الرياضي المستخدم في كل طريقة ، كما يلاحظ أيضا أن قيمة المنوال بأي من الطريقتين أقل من قيمة الوسيط أقل من قيمة الوسط الحسابي كما هو واضح من شكل منحنى التوزيع في المطلوب السابق مباشرة .

(جـ) بالرسم من المدرج التكراري :

يتم حساب قيمة المنوال بالرسم من المدرج التكراري ويكتفى في هذه الحالة برسم المستطيلات التي تمثل الفئة المنوالية والسابقة واللاحقة لها كما يتضح من الشكل الاتى :



ومن الرسم السابق نجد أن قيمة المنوال تساوى تقريبا ٧٤ جنيه .

ويلاحظ أن قيمة المنوال يجب ألا تتعدى الفئة المنوالية وهى (٦٥– ٨٥) فى مثالنا هذا .

٤– حساب قيمة الوسط الهندسى (هـ) :

نعلم أن قيمة الوسط الهندسى – فى حالة البيانات المبوبة – نحصل عليه من العلاقة التالية :

$$\text{لو هـ} = \frac{\text{مجـ ك}_{\text{ر}}\text{لوس}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}}$$

وبالكشف فى جداول الاعداد المقابلة

للوغاريتمات نحصل على قيمة الوسط الهندسي (هـ) ، ولحساب قيمة (هـ)

من العلاقة السابقة لابد من تكوين الجدول التالى :

فئات الدخل (س)	عدد الاسر (التكرار = $ك_ر$)	مراكز الفئات ($س_ر$)	لوغاريتم مراكز الفئات (لو $س_ر$)	$ك_ر$ × لو $س_ر$
١٥-	٤	١٥	١.١٧٦١	٤.٧٠٤٤
٢٥-	٦	٣٥	١.٥٤٤١	٩.٢٦٤٦
٤٥-	١٥	٥٥	١.٧٤٠٤	٢٦.١٠٦٠
٦٥-	٢٢	٧٥	١,٨٧٥١	٤١,٢٥٢٢
٨٥-	١٣	٩٥	١.٩٧٧٧	٢٥.٧١٠١
١٠٥-	٧	١١٥	٢,٠٦٠٧	١٤.٤٢٤٩
١٢٥-	٥	١٣٥	٢,١٣٠٣	١٠,٦٥١٥
١٤٥-١٦٥	٣	١٥٥	٢.١٩٠٣	٦,٥٧٠٩
المجموع	٧٥	-	-	١٣٨.٦٨٤٦

$$لو\ هـ = \frac{مجـ\ ك_ر لو س_ر}{مجـ\ ك_ر}$$

$$= \frac{١٣٨,٦٨٤٦}{٧٥} = ١,٨٤٩١$$

وبالكشف فى جداول الاعداد المقابلة للوغاريتمات نجد أن :

الوسط الهندسى (هـ) = ٧٠,٦٥ جنيها

<u>٥- حساب قيمة الوسط التوافقى :</u>

يتم حساب قيمة الوسط التوافقى للبيانات المبوبة من العلاقة :

$$ق = \frac{\text{مجـ } ك_ر}{\text{مجـ } \left(\frac{ك_ر}{س_ر} \right)}$$

ولحساب قيمة (ق) من العلاقة السابقة ننشأ الجدول التالى :

فئات الدخل	عدد الاسر (التكرار = ك)	مراكز الفئات (س ر)	(ك ر / س ر)
٥-	٤	١٥	٠.٢٦٦٧
٢٥-	٦	٣٥	٠,١٧١٤
٤٥-	١٥	٥٥	٠,٢٧٢٧
٦٥-	٢٢	٧٥	٠.٢٩٣٣
٨٥-	١٣	٩٥	٠.١٣٦٨
١٠٥-	٧	١١٥	٠.٠٦٠٩
١٢٥-	٥	١٣٥	٠.٠٣٧٠
١٤٥-١٦٥	٣	١٥٥	٠.٠١٩٤
المجموع	٧٥	-	١.٢٥٨٢

$$ق = \frac{\text{مجـ } ك_ر}{\text{مجـ } \left(\frac{ك_ر}{س_ر} \right)}$$

$$= \frac{٧٥}{١,٢٥٨٢} = ٥٩,٦١ \text{ جنيها}$$



الباب الرابع

مقاييس التشتت

Measures of Dispersion

لقد تكلمنا فى الباب السابق عن المقاييس المختلفة للنزعة المركزية – المتوسطات – ورأينا أن هذه المقاييس تعتبر من أهم خصائص التوزيع ، حتى أن المجتمعات المختلفة تتميز عن بعضها بقيم هذه المقاييس ، فنجد مثلا أن متوسط أعمار الرجال فى المدن يختلف عنه فى القرى ، وتسمى هذه المقاييس الاحصائية بمعالم المجتمع .

ومن هذا تبين لنا أن المتوسطات (الوسط الحسابى – الوسيط – المنوال – ...) تعطينا فكرة عن التوزيع التكرارى ، الا أن هذه الفكرة لا تكون كاملة إذ أن المتوسطات وحدها لا تكفى لاعطاء فكرة دقيقة عن المجموعة موضوع الدراسة ، حيث لا نستطيع تبين طبيعتها ولا كيفية توزيع مفرداتها ، كما أن استخدام المتوسطات فقط لمقارنة عدة مجموعات لا يكفى لإظهار حقيقة المقارنة فقد تتساوى القيمة المتوسطة لمجموعتين ممن البيانات بينما تختلف المجموعتيين عن بعضهما كل الأختلاف ، فقد تكون مفردات إحدى المجموعتين متقاربة بعضها من بعض (أى تتركز قيم المجموعة حول متوسطها) أو متبعثرة

(مشتتة) . وعلى ذلك فلكى نتمكن من وصف ومقارنة مفردات مجموعتين ، بدقة يجب ألا نقتصر على مقارنة متوسطى المجموعتين بل يجب أن نصف درجة اختلاف مفردات كل من المجموعتين بعضهما عن بعض أو عن متوسطاتهما أو بعبارة أخرى وصف درجة تشتتها وهذا ما يتم عن طريق قياس التشتت (درجة التجانس) لكل مجموعة على حده .

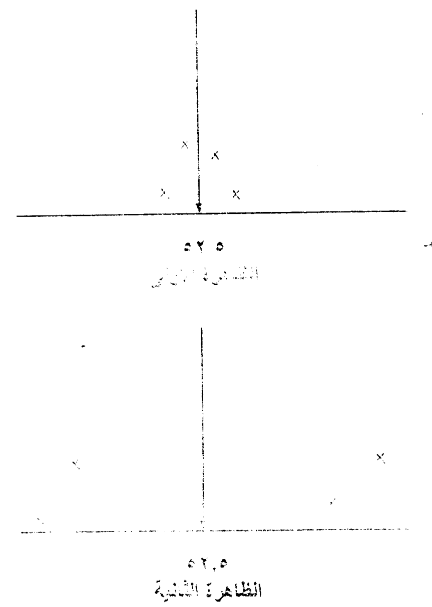
[illegible]

نفرض أن لدينا مجموعتين من الطلبة وكانت درجات المجموعة الاولى ٥٠ ، ٥٢ ، ٥٥ ، ٥٣ بمتوسط ٥٢,٥ درجة وكانت درجات المجموعة الثانية ١٠ ، ٩٠ ، ٨٠ ، ٣٠ بمتوسط ٥٢,٥ درجة أيضا فإذا ما اكتفينا بمقارنة الوسطين الحسابيين للمجموعتين (كل منهم يساوى ٥٢,٥) - فإننا سوف نستنتج أن مستوى الاستيعاب لأفراد المجموعتين متساوى أيضا وهذا غير صحيح لانهما مختلفان فعلا ، فبينما نجد أن درجات المجموعة الأولى محصورة بين ٥٠ ، ٥٥ درجة (أى أن الفرق بين اعلى درجة وأقل درجة هو ٥ درجات) نجد أن اختلاف الدرجات للمجموعة الثانية (تشتت الدرجات) كبيرا فالفرق بين أكبر الدرجات وأصغرها يساوى ٨٠ درجة . وهذا ما يجعلنا نستنتج أن قيم المجموعة الاولى مركزة (متجانسة) بينما قيم المجموعة الثانية مبعثرة (مشتتة) .

وبصفة عامة ، فيقصد بتشتت أى مجموعة من المفردات مدى التباعد أو الاختلاف بين قيم هذه المفردات ، ومن البديهى أنه أذا كان هذا التشتت صغيرا فهذا يدل على أن مقدار الاختلاف بين هذه القيم كان قليلا (كما هو واضح من الشق الاول من المثال السابق) حيث لوحظ أن معظم القيم تقترب من قيمتها المتوسطة (فإذا تساوت جميع قيم الظاهرة فإن التشتت فى هذه الحالة يساوى صفراً) إما إذا كان التشتت كبيرا فإن هذا يعنى أن الأختلاف بين قيم مفردات الظاهرة كبيرا (أى تنتشر مفردات الظاهرة مبتعدة عن قيمتها المتوسطة)

وعموما فكما يعنينا تقدير أو حساب القيمة المتوسطة للظاهرة محل الدراسة ، فإنه لا يقل عنه أهمية تقدير أو قياس درجة تجانس أو تشتت مفردات الظاهرة ، لذلك سوف تستخدم معايير كمية لقياس المدى التى تنتشر عليه البيانات أو لقياس درجة تبعثر القيم أو التفافها حول قيمتها المتوسطة – حيث تعرف هذه المقاييس بمقاييس التشتت .

فإذا حاولنا تمثيل مفردات ظاهرتى الدرجات فى المثال السابق بيانيا ، نجد أن قيم الظاهرة الاولى (درجات الطلاب) أكثر تجانس (أقل تشتت) بينما يلاحظ أن قيم الظاهرة الثانية أقل تجانس (أى أكثر تشتتا – أكثر تباين) .



٥٢ م

الظاهرة الأولى

٥٢،٥ م

الظاهرة الثانية

ومما يجب ملاحظته أنه إذا كان مقدار التشتت صغيرا فإن القيمة المتوسطة تمثل [illegible] جيدا أو [illegible] على العكس من ذلك إذا كان مقدار التشتت كبيرا فإن القيمة المتوسطة في هذه الحالة [illegible] تعبيرا [illegible] التمثيل الظاهرة موضوع الدراسة .

(٤-١) مقاييس التشتت المطلقة :-

عموما سوف نستعرض مجموعة من المقاييس الاحصائية التى تمكننا من قياس درجة تشتت بيانات أى مجموعة والتى تختلف فيما بينها من حيث دقتها وطرق حسابها وأهم هذه المقاييس :

١) المدى

(٢) نصف المدى الربيعى

(٣) الانحراف المتوسط

(٤) الانحراف المعيارى والتباين

وسوف نتعرف على كيفية حساب أى من هذه المقاييس فى حالتى البيانات الغير مبوبة (الخام) والبيانات المبوبة (التوزيعات التكرارية) .

(٤-١-١) المدى *Range* :-

يعد المدى من أبسط مقاييس التشتت واسهلها فى الحساب ، الامر الذى جعله شائع الاستخدام فى كثير من الميادين ويعرف بأنه الفرق بين أكبر وأصغر قراءة فى المجموعة ، فإذا كان المدى صغيرا تكون المجموعة متقاربة (متجانسة) وعلى العكس إذا كان المدى كبيرا فأنه يدل على أن مفردات المجموعة مبعثرة أو مشتته أو متباعدة عن بعضها .

من اصل المصدر

<u>أولاً: المدى في حالة البيانات غير المبوبة :-</u>

فمثلاً إذا كان لدينا القيم س١ ، س٢ ، س٣ ، ... ، سن مرتبة تصاعدياً بحسب قيمتها ، بحيث تكون (س١) هي أصغر القيم ، (سن) هي أكبر القيم فإن المدى لهذه المجموعة يعبر عنه كالآتي :

[illegible]

في مثال (١) يلاحظ أن :-

المدى للمجموعة الأولى من الدرجات = ٥٥ − ٥٠ = ٥ درجة بينما

[illegible] المجموعة الثانية من الدرجات = ٩٠ − ١٠ = ٨٠ درجة

[illegible] بأن درجات المجموعة الأولى [illegible]

[illegible] المجموعة الثانية .

<u>ثانياً: المدى في حالة البيانات المبوبة :-</u>

[illegible] المدى في حالة البيانات المبوبة (في جدول تكراري [illegible]

[illegible] :

المدى = الحد الأعلى للفئة الأخيرة − الحد الأدنى للفئة الأولى

أى

> المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

<u>مثال (٢) : –</u>

أحسب المدى للتوزيع التكرارى للانفاق الشهرى للاسر (المثال العام المحلول – الباب الثالث)

المدى = ١٦٥ – ٥ = ١٦٠

<u>* ملاحظات عامة على المدى :</u>

مما سبق يمكن للقارىء أن يلاحظ على المدى ما يلى :

- أنه مقياس ترتيبى سهل سريع الحساب إذ أنه يستخدم مفردتيين اثنتين فقط تقع أولاهما عند نهاية التوزيع والأخرى عند بدايته.
- وبسبب ذلك فإنه لا يمكن حسابه من توزيع تكرارى مفتوح كما أن اقتصاره على هاتين المفردتين يعنى اهماله لقيمة باقى المعلومات المتاحة عن الظاهرة .

• ونظرا لأعتماد المدى على المفردتين من حيث الموقع فانه مقياس غير مستقر ويتأثر سريعا بتلك القيم المتطرفة مما يؤدى الى اعطانه لصوره غير صحيحة عن درجة التجانس وأيضا توزيع الظاهرة .

• إضافة إلى ما سبق فانه بزيادة عدد المفردات تزداد فرص ظهور القيم المتطرفة منها فى القيمة ولو أن هذا المقياس جيد – من الناحية الاحصائية – لعكس تجانس أكبر بازدياد عدد المفردات ، فمن المعروف أنه كلما ذاد حجم العينة التى تدرس عن طريقها الظاهرة كلما زاد تجانس توزيع تلك الظاهرة .

ولكن برغم عيوبه هذه إلا أنه شائع الاستخدام فى اسلوب ضبط الانتاج وكذلك فى دراسة التغيرات فى درجات الحرارة اليومية . خاصة أنه عادة ما تستخدم فى أسلوب ضبط الانتاج عينات صغيرة ومتساوية الحجم ، وعليه فإن حجم العينة لن يؤثر على المدى فيعيبه كمقياس جيد لدرجة التجانس .

كما يستخدم المدى ايضا كوسيلة سريعة للتحقق من صحة العمليات الحسابية الخاصة بتقدير مقاييس أخرى للتشتت إذا أنه من المتوقع ألا يزيد أو يقل الانحراف المعيارى عن ٧ أمثال المدى ، وهذه العلاقة التجريبية ليست دليلا قاطعا على صحة أو عدم صحة القيمة المحسوبة للانحراف المعيارى – فعدم تحققها يعنى احتمال وجود خطأ حسابى .

(٤-١-٣) نصف المدى الربيعى Semi Interquartile Range :-

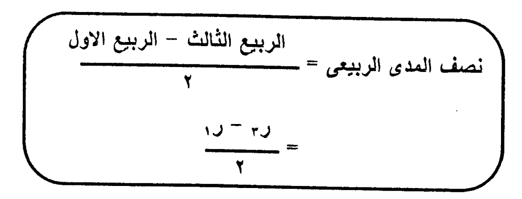
سبق أن رأينا أن المدى كمقياس للتشتت شديد التاثر بالقيم المتطرفة ، وقد دعا ذلك إلى محاولة معالجة حساسية للقيم المتطرفة باستبعادها من طرفى التوزيع فإذا ما أستبعدت المفرده الاولى ($س_١$) والخيرة ($س_ن$) فان الفرق بين ($س_{ن-١}$) ، $س_٢$ يمكن أن يعتبر مقياسا أفضل للتشتت وهو ما أطلق عليه اسم المدى الأول .

أى أن :

$$\text{المدى الاول} = س_{ن-١} - س_٢$$

وبنفس الطريقة يمكن استبعاد مفردتين من نهاية التوزيع (بعد ترتيب المفردات) ومفردتين عند بدايته فنحصل على المدى الثانى وهكذا. وتعرف هذه المقاييس بشبيهات المدى .

واعتمادا على نفس الفكرة فقد استنبط الاحصائيون مقياس يعرف بنصف المدى الربيعى وذلك بأخذ نصف الفرق بين الربيعين الثالث والأول وهو ما يعرف أيضا باسم الانحراف الربيعى حيث :



$$\text{نصف المدى الربيعى} = \frac{\text{الربيع الثالث} - \text{الربيع الاول}}{2} = \frac{ر_3 - ر_1}{2}$$

ويتميز نصف المدى الربيعى بأنـه سـهل الحسـاب والتطبيق ، كمـا أنـه لا يتأثر الى حد ما بالقيم المتطرفة بالاضافـة إلـى أنـه المقيـاس الوحيـد مـن مقـاييس التشتت المطلقة الذى يمكن حسابه من الجداول التكرارية المفتوحـة والمغلقـة ، إلا أنه يعاب عليه أنه يعتمد فى حسابه على قيمتين فقط همـا ($ر_1$ ، $ر_3$) ويهمـل بـاقى القيم شأنه فى ذلك شأن المدى أيضـا ، فضـلا عـن أنـه يتـأثر بحجـم العينـة كشـأن المدى ، فضلا عن أنه لا يخضع خضوعا تامـا للعمليـات الحسـابية إذ أنـه لا يمكـن حساب نصف المدى الربيعى العام لعدة مجموعات جزئيـة وخاصـة إذا علـم نصـف المدى الربيعى لكل مجموعة جزئية .

وتتوقـف جـودة نصـف المـدى الربيعـى علـى درجـة تركـيز البيانـات عنـد الربيعين الادنى والأعلى ، فإذا كان هناك فراغ فى البيانـات حـول الربيعيـن يصبـح نصف المدى الربيعى غير ملائم لقياس تشتت الظاهرة .

<u>مثال (٣) :-</u>

أوجد نصف المدى الربيعى لتوزيع الانفاق الشهرى (المثال العام المحلول - الباب الثالث)

<u>الحل</u>

لوحظ أن قيم $ر_٣$ = ٩٩,٢٣   **ولأن** $ر_١$ = ٥٦,٦٧

$$\therefore \text{نصف المدى الربيعى} = \frac{ر_٣ - ر_١}{٢}$$

$$= \frac{٩٩,٢٣ - ٥٦,٦٧}{٢} = ٢١,٢٨$$

(٤-١-٣) الانحراف المتوسط *Mean Deviation*:-

سبق أن استخدمنا المدى كمقياس للتشتت على اساس تباعد القيم عن بعضها ، ومن الواضح أنه لو كانت القيم قريبة من بعضها فإنها تكون مركزة (متجمعة) حول قيمتها المتوسطة ، وكلما كانت مبعثرة كلما تباعدت القيم عن هذه القيمة (أى كلما ازداد تباعدها عن هذه القيمة) وعلى ذلك فمن الممكن أن نستحدث مقياسا آخر للتشتت على اساس الفروق بين القيم المختلفة وقيمته المتوسطة - مثل الوسط الحسابى أو الوسيط .... الخ .

ولكننا نعلم أن مجموع انحرافات القيم عن وسطها الحسابى = صفرا (لأن مجموع الانحرافات الموجبة عن وسطها الحسابى = مجموع الانحرفات السالبة

عنه) ولهذا فلا نستطيع أخذ مجموع الانحرفات عن الوسط الحسابى كمقياس للتشتت ، إلا أنه فى الامكان أن نهمل اشارات الانحرافات ونأخذ مجموع الانحرافات باهمال الاشارة . ولما كان مجموع الانحرافات يتوقف على عدد القيم فيكون المقياس المقترح هو مجموع الانحرافات عن القيمة المتوسطة مع اهمال الاشارة مقسوما على عدد القيم وهذا ما نسميه بالأنحراف المتوسط .

أولا : الانحراف المتوسط فى حالة البيانات غير المبوبة :-

إذا كان لدينا مجموعة من القيم $س_1$ ، $س_2$ ، .... $س_ن$ فإن الانحراف المتوسط عن الوسط الحسابى يعرف كالاتى :

$$\text{الانحراف المتوسط عن الوسط الحسابى} = \frac{\sum_{ر=1}^{ن} |س_ر - \bar{س}|}{ن}$$

حيث يرمز الخطان الرأسيان الى أن الانحرافات تؤخذ بدون الاشارة الجبرية (أى الانحرافات المطلقة) .

كما أن الانحراف المتوسط يمكن أن يكون مأخوذا عن الوسيط حيث :

$$\text{الانحراف المتوسط عن الوسيط} = \frac{\sum_{ر=1}^{ن} |س_ر - ر_2|}{ن}$$

ويجدر الاشارة الى أنه يمكن حساب الانحراف المتوسط عن أى قيمة متوسطة أخرى ، كالمنوال أو الوسط الهندسى ... الخ . الا أن هذا نادر الحدوث . والمقياس المستخدم غالبا هو الانحراف المتوسط عن الوسط الحسابى ، وعموما إذا ذكر الانحراف المتوسط دون تخصيص فى التمرين فيكون المقصود هو أخذ الانحرافات المطلقة عن الوسط الحسابى .

مثال (٤) :-

احسب الانحراف المتوسط عن كل من الوسط الحسابى والوسيط لمجموعة القيم الاتية :

٣٠ ، ٣٣ ، ٣٥ ، ٣٧ ، ٤٠

الحل

- **لحساب الانحراف المتوسط عن الوسط الحسابى ، فلابد أولا من حساب الوسط الحسابى ($\overline{س}$) لمجموعة القيم السابقة حيث :**

$$\overline{س} = \frac{مجـ\ س_ر}{ن} = \frac{١٧٥}{٥} = ٣٥$$

ويكون مجموع الانحرافات المطلقة للقيم عن وسطها الحسابى كما هو واضح من الجدول التالى:

| سر - س̄ |	س
| ٣٠ - ٣٥ | = ٥	٣٠
| ٣٣ - ٣٥ | = ٢	٣٣
| ٣٥ - ٣٥ | = صفر	٣٥
| ٣٧ - ٣٥ | = ٢	٣٧
| ٤٠ - ٣٥ | = ٥	٤٠
١٤	المجموع

∴ الانحراف المتوسط عن الوسط = $\frac{\text{مجـ } |س_ر - \bar{س}|}{ن}$

$= \frac{١٤}{٥} = ٢,٨$

• لحساب الانحراف المتوسط عن الوسيط ، فلابد أولا من حساب الوسيط ($ر_٢$) لمجموعة القيم السابقة حيث :

$ر_٢ = ٣٥$ (ونظرا لتساوى قيمتى $\bar{س}$ ، $ر_٢$ ، وبالتالى توقع تساوى النتائج فى الحالتين بعد تطبيق نفس الخطوات فى حالة الوسط الحسابى .

∴ الانحراف المتوسط عن الوسيط = $\frac{\text{مجـ } |س_ر - ر_٢|}{ن}$

$= \frac{١٤}{٥} = ٢,٨$

ثانيا : الانحراف المتوسط فى حالة البيانات المبوبة :-

. يحسب الانحراف المتوسط فى حالة البيانات المبوبة (التوزيعات التكرارية)

على النحو التالى :

$$\text{الانحراف المتوسط عن الوسط الحسابى} = \frac{\text{مجـ}\ |\text{س}_\text{ر} - \overline{\text{س}}|\ \text{ك}_\text{ر}}{\text{مجـ}\ \text{ك}_\text{ر}}$$

كما أن :

$$\text{الانحراف المتوسط عن الوسيط} = \frac{\text{مجـ}\ |\text{س}_\text{ر} - \text{ر}_\text{٢}|\ \text{ك}_\text{ر}}{\text{مجـ}\ \text{ك}_\text{ر}}$$

حيث $\text{س}_\text{ر}$ تشير الى مراكز الفئات.

مثال(٥) :-

المطلوب حساب الانحراف المتوسط عن الوسط الحسابى ثم عن الوسيط للتوزيع التكرارى التالى :

فئات الوزن	٣٠ -	٤٠ -	٤٨ -	٥٤ -	٦٠ - ٨٠	المجموع
عدد الأشخاص	١٠	١٥	٥٠	٢٢	٣	١٠٠

الحل :

لحساب الانحراف المتوسط عن الوسط والوسيط علينا اولا حساب كل من

الوسط والوسيط كما هو موضح فى الجدول التالى :

فئات الوزن	عدد الاشخاص (كر)	سر	سر كر	|سر - س̄|	|سر - س̄| كر	حدود عليا للفئات	تكرار متجمع صاعد	|سر - ر.|	|سر - ر.| كر
٣٠ -	١٠	٣٥	٣٥٠	١٥,٢٤	١٥٢,٤٠	أقل من ٤٠	١٠	١٦	١٦٠
٤٠ -	١٥	٤٤	٦٦٠	٦,٢٤	٩٣,٦٠	أقل من٤٨	٢٥	٧	١٠٥
٤٨ -	٥٠	٥١	٢٥٥٠	٠,٧٦	٣٨,٠٠	أقل من ٥٤	٧٥	صفر	صفر
٥٤ -	٢٢	٥٧	١٢٥٤	٦,٧٦	١٤٨,٧٢	أقل من ٦٠	٩٧	٦	١٣٢
٦٠- ٨٠	٣	٧٠	٢١٠	١٩,٧٦	٥٩,٢٨	أقل من ٨٠	١٠٠	١٩	٥٧
المجموع	١٠٠	—	٥٠٢٤	—	٤٩٢	—	—	—	٤٥٤

١- إيجاد الانحراف المتوسط عن الوسط الحسابى :

من الجدول السابق نلاحظ أن :

$$\bar{س} = \frac{\text{مجـ سر كر}}{\text{مجـ كر}} = \frac{٥٠٢٤}{١٠٠} = ٥٠,٢٤٥ \text{ كيلو جرام}$$

$$\therefore \text{الانحراف المتوسط عن الوسط الحسابى} = \frac{\text{مجـ } |\text{سر} - \bar{س}|\text{ كر}}{\text{مجـ كر}}$$

$$= \frac{٤٩٢}{١٠٠} = ٤,٩٢$$

٢- إيجاد الانحراف المتوسط عن الوسيط :

$$\text{ترتيب } ر_٢ = \frac{\text{مجـ ك}}{٢} = \frac{١٠٠}{٢} = ٥٠$$

$$\therefore \quad ر_٢ = ف\, د_٢ + \frac{\frac{\text{مجـ ك}}{٢} - ك\, د_٢}{ك\, أ_٢ - ك\, د_٢} \times ل_٢$$

$$= ٤٨ + \frac{٥٠ - ٢٥}{٧٥ - ٢٥} \times ٦$$

= ٤٨ + ٣ = ٥١ كيلو جرام

$$\therefore \text{الانحراف المتوسط عن الوسيط} = \frac{\text{مجـ } | س_ر - ر_٢ | ك_ر}{\text{مجـ } ك_ر}$$

$$= \frac{٤٥٤}{١٠٠} = ٤,٥٤$$

وعموما يلاحظ أن الانحراف المتوسط سواء عن الوسط أو الوسيط أكثر دقة من المدى ونصف المدى الربيعى حيث أنه لا يتأثر بالقيم الشاذة كما أنه يعتمد فى حسابه على جميع المفردات ، إلا أنه يعاب عليه الصعوبة فى الحساب فضلا عن أن حسابه يتطلب إهمال الاشارة وذلك للتغلب على مشكلة هامة وهى أن مجموع إنحرافات القيم عن الوسط الحسابى (أو عن الوسيط فى حالة التوزيعات المتماثلة) تساوى صفرا وهذا الاهمال ليس له ما يبرره رياضيا ويجعل المقياس غير قابل للاستخدام فى معالجة جبرية أخرى . كما أنه لا يمكن استخدامه فى حالة

التوزيعات التكرارية المفتوحة وذلك لاعتماده على مراكز الفئات التى يمكن حسابها فى مثل هذه التوزيعات .

ولكن هذه المقاييس الثلاثة لقياس التشتت المطلق لأى ظاهرة تعد قليلة الاستخدام فى الواقع العملى نظرا لوجود مقياس أفضل يسهل معالجته حسابيا . علاوة على توافر خصائص التقدير الجيد لهذا المقياس والذى يطلق عليه الاحصائيون الانحراف المعيارى (الجذر التربيعى للتباين) وهو ما سندرسه فى المرحلة التالية من هذا الباب .

(٤-١-٤) التباين والانحراف المعيارى

Variance & Standard Deviation

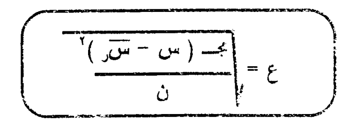
من المعروف أنه كلما ارتفعت درجة تجانس التوزيع كلما اقتربت المشاهدات المختلفة من وسطها الحسابى ، والعكس صحيح أى أنه كلما زادت درجة عدم التجانس كلما زاد انتشار المفردات وتباعدها عن وسطها الحسابى . لذلك فانه فى مقدورنا استخدام هذه الخاصية فى تعريف مقياس لدرجة التشتت أو الانتشار أو عدم التجانس . ولكننا نعلم أن مجموع انحرافات القيم عن وسطها الحسابى يساوى صفرا وعليه فاننا لا نستطيع استخدام هذا المجموع كاساس لقياس التشتت مالم نعمل بوسيلة أو بأخرى تلاشى المجموع ، ونستطيع أن نعالج مشكلة تلاشى مجموع فروق القيم عن وسطها وذلك بتربيع تلك الفروق . فـاذا ما

حساب متوسط مجموع مربعات الفروق بين القيم المختلفة ووسطها الحسابى لكان الناتج هو المقياس المعروف بالتباين وجذره التربيعى الموجب هو مقياس التشتت المعروف باسم الانحراف المعيارى .

ويعتبر الانحراف المعيارى من أهم مقاييس التشتت المطلقة استخداما فهو معرف تعريفا دقيقا ويدخل فى حسابه كل المفردات ويخضع كذلك للطرق الرياضية فهو يستخدم على نطاق وساع للغاية فى نظرية التقديرات وكذلك فى اختبارات الفروق الاحصائية .

أولا : الانحراف المعيارى (ع) فى حالة البيانات غير المبوبة :-

إذا كان لدينا مجموعة من البيانات الغير مبوبة $س_1$ ، $س_2$ ، ... ، $س_ن$ فإن الانحراف المعيارى يعرف كالاتى :



$$ع = \sqrt{\frac{مجـ (س - \bar{س})^2}{ن}}$$

يتضح من التعريف السابق أننا قد تخلصنا من العيب الذى تعرض له الانحراف المتوسط حيث تم التخلص من اشارة الانحراف عن المتوسط بطريقة

جبرية وذلك بتربيع تلك الانحرافات ثم أخذ الجذر التربيعي الموجب لمتوسط مربعات الانحرافات .

مثال (٦) :-

احسب الانحراف المعيارى لمجموعة القراءات التالية :

٤٠ ، ٤٤ ، ٢٨ ، ٣٦ ، ٣٢

الحل :-

$$\text{الوسط الحسابى } (\bar{س}) = \frac{\text{مجـ } س}{ن} = ٣٦$$

ولايجاد الانحراف المعيارى نكون الجدول التالى :

س	(س - $\bar{س}$)	(س - $\bar{س}$)^2
٤٠	٤	١٦
٤٤	٨	٦٤
٢٨	٨-	٦٤
٣٦	صفر	صفر
٣٢	٤ -	١٦
المجموع	-	١٦٠

$\therefore$ الانحراف المعيارى (ع) = $\sqrt{\frac{1}{ن} \text{مجـ} (س - \bar{س})^2}$

$= \sqrt{\frac{1}{5} \times 160}$

$= \sqrt{32} = 5{,}656$

يلاحظ أنه لحساب الانحراف المعيارى باستخدام الصيغة السابقة فاننا نمر بعده عمليات حسابية ولذلك توجد صيغة أفضل من الصيغة السابقة يمكن ايجادها كالاتى :

$\text{مجـ} (س - \bar{س})^2 = \text{مجـ} (س^2 + \bar{س}^2 - 2س\bar{س})$

$= \text{مجـ} س^2 + ن\bar{س}^2 - 2\bar{س} \text{ مجـ} س$

$= \text{مجـ} س^2 - ن\bar{س}^2$

$= \text{مجـ} س^2 - \frac{(\text{مجـ} س)^2}{ن}$

واحيانا يفضل ايجاد الانحراف المعيارى استخدام الصيغة التالية:-

$$ع = \sqrt{\frac{1}{ن} \left(\text{مجـ} س^2 - \frac{(\text{مجـ} س)^2}{ن} \right)}$$

ويطلق على المقدار الموجود تحت الجذر فى الصيغة السابقة اسم التباين ($ع^{٢}$) ، ولذلك فالتباين يعرف بأنه متوسط مربعات انحرافات القيم عن وسطها الحسابى ولا يعتبر مقياسا للتشتت .

مثال (٧) :-

احسب الانحراف المعيارى لقيم المثال السابق باستخدام الصيغة السابقة (صيغة مجاميع القيم) :

الحل :

س	$س^{٢}$
٤٠	١٦٠٠
٤٤	١٩٣٦
٢٨	٧٨٤
٣٦	١٢٩٦
٣٢	١٠٢٤
١٨٠	٦٦٤٠

$$\therefore ع = \sqrt{\frac{١}{ن}\left(مجـ\, س^{٢} - \frac{(مجـ\, س)^{٢}}{ن} \right)}$$

$$= \sqrt{\frac{١}{٥}\left(٦٦٤٠ - ٦٤٨٠ \right)}$$

$= \sqrt{٣٢} = ٥,٦٥٦$

وهو نفس الناتج باستخدام الصيغة الاصلية .

وللانحراف المعيارى مجموعة من الخصائص الهامة نذكر منها ما يلى :-

١ - لا يتأثر الانحراف المعيـارى بـالطرح ولا بـالجمع ويمكـن اثبـات ذلـك رياضيـا على النحو التالى :

نفرض أن لدينا مجموعة من القراءات هى :

$س_١$ $س_٢$ ، $س_ن$ بوسط حسابى ($\overline{س}$) = $\frac{مجـ س}{ن}$.

والانحراف معيارى (ع) = $\sqrt{\frac{١}{ن} مجـ (س - \overline{س})^٢}$

وبفرض أننا أضفنا المقدار الثابت (أ) على كل قراءة من القراءات لتصبـح

($س_١$ + أ) ، ($س_٢$ + أ) ، .... ($س_ن$ + أ)

ويكون الوسط الحسابى للقيم الجديدة عبارة عن :

$\overline{\overline{س}} = \frac{مجـ س}{ن} + أ$

$= \overline{س} + أ$

والانحراف المعيارى عبارة عن :

$ع' = \sqrt{\frac{١}{ن} مجـ (س + أ - \overline{\overline{س}})^٢}$

$$= \sqrt{\frac{١}{ن} \text{مجـ} (س + أ - \overline{س} - أ)^{٢}}$$

$$= \sqrt{\frac{١}{ن} \text{مجـ} (س - س)^{٢}} = ع$$

معنى ذلك أن الانحراف المعيارى لا يتأثر بالجمع ولا بالطرح

٢ - يتأثر الانحراف المعيارى بالضرب والقسمة ويمكن اثبات ذلك رياضيا كالاتى :

نفرض أن قيم المتغير س هى $س_{١}$ ، $س_{٢}$ ، ... ، $س_{ن}$

بمتوسط $\overline{س} = \frac{\text{مجـ } س}{ن}$ **وانحراف معيارى** $ع = \sqrt{\frac{١}{ن} \text{مجـ} (س - \overline{س})^{٢}}$

وبضرب كل قيمة من قيم المتغير س فى الثابت (أ) فتصبح القيم أ $س_{١}$ ، أ $س_{٢}$ ، .... ، أ $س_{ن}$ بمتوسط $\overline{\overline{س}}$ = أ $\overline{س}$

ويحسب الانحراف المعيارى للقيم الجديدة كالاتى :

$$ع^{/} = \sqrt{\frac{١}{ن} = (أ س - \overline{\overline{س}})^{٢}}$$

$$= \sqrt{\frac{١}{ن} \text{مجـ} (أ س - أ \overline{س})^{٢}}$$

$$= \sqrt{\frac{\text{أ}^{2}}{\text{ن}} \sum (\text{س} - \bar{\text{س}})^{2}} = \text{أ}\ \text{ع}$$

ويعنى ذلك أن الانحراف المعيارى للقيـم بعد الضـرب فـى أى مقـدار ثـابت عبارة عن الانحراف المعيارى لهذه القيـم مضروبـا فـى الثـابت (لاحـظ أنـه يمكـن اثبات هذه الخاصية فى حالة القسمة باتباع نفس الاسلوب الرياضى) .

٣ - مجموع مربعات انحرافات القيم عن وسطها الحسابى أقل ما يمكن أى ان :

$$\sum (\text{س} - \bar{\text{س}})^{2} < \sum (\text{س} - \text{أ})^{2} \quad ، \ \text{أ} \neq \bar{\text{س}}$$

ولاثبات ذلك نظريا نفرض أن لدينا مقدار ثابت أ .

$$\therefore \sum (\text{س} - \text{أ})^{2} = \sum [(\text{س} - \bar{\text{س}}) + (\bar{\text{س}} - \text{أ})]^{2}$$

$$= \sum [(\text{س} - \bar{\text{س}})^{2} + 2 (\text{س} - \bar{\text{س}}) (\bar{\text{س}} - \text{أ}) + (\bar{\text{س}} - \text{أ})^{2}]$$

$$= \sum (\text{س} - \bar{\text{س}})^{2} + 2 (\bar{\text{س}} - \text{أ}) \sum (\text{س} - \bar{\text{س}}) + \text{ن}(\bar{\text{س}} - \text{أ})^{2})$$

وحيث أن $\sum (\text{س} - \bar{\text{س}})$ = صفر (من الخصائص السابقة للوسط الحسابى)

$$\therefore \quad \sum (\text{س} - \text{أ})^{2} = \sum (\text{س} - \bar{\text{س}})^{2} + \text{ن} (\bar{\text{س}} - \text{أ})^{2}$$

يلاحظ فى المعادلة السابقة أن االطرف الايمن للمعادلـة [$\sum (\text{س} - \text{أ})^{2}$] يســاوى حاصــل جمــع مقدارين أحدهمــا = [$\sum (\text{س} - \bar{\text{س}})$] والأخـر [$\text{ن} (\bar{\text{س}} - \text{أ})^{2}$] وهو مقدار موجب دائما [حيـث يتكـون مـن (ن) الموجبـة ومجموع مربعات الفرق بين $\bar{\text{س}}$ ، أ وهذه الفروق موجبة أيضا] وبالتالى فان :

$$\sum (\text{س} - \bar{\text{س}})^{2} < \sum (\text{س} - \text{أ})^{2}$$

كما يمكننا حساب الانحراف المعيارى باستخدام إحدى طريقتين (طريقة الفروق البسيطة - طريقة الفروق المعدلة) كما هو الحال عند حساب الوسط الحسابى ($\bar{س}$) .

طريقة الفروق البسيطة :-

تهدف هذه الطريقة الى اختصار العمليات الحسابية وخاصة عندما يكون عدد المفردات كبيرا ، الامر الذى يؤدى الى تحقيق وفرة فى الجهد الحسابى علاوة على تقليل احتمالات الخطأ .

فإذا اعتبرنا أن ح = (سر - أ) تمثل الفروق البسيطة من قيم سر وأى وسط فرضى (أ) فإن :

$$ع = \sqrt{\frac{مجـ ح^{٢}}{ن} - \left(\frac{مجـ ح}{ن}\right)^{٢}}$$

طريقة الفروق المعدلة :-

وتقضى هذه الطريقة من طرق حساب الانحراف المعيارى بقسمة جميع قيم ($ح$) على مقدار ثابت ($ب$) بدون باقى ومن ثم نحصل على الفروق المعدلة ($ح'$)

حيث $ح' = \frac{ح}{ب}$ وبالتالى فإن :

$$ع = ب\sqrt{\frac{\text{مجـ}\ ح'^{2}}{ن} - \left(\frac{\text{مجـ}\ ح'}{ن}\right)^{2}}$$

وذلك لأن الانحراف المعيارى كما علمنا من الخاصية الثانية أنه يتأثر بعملية القسمة .

مثال(٨)

احسب الانحراف المعيارى للقيم الأتيـة : ٦٢٥ ، ٧٠٠ ، ٦٥٠ ، ٧٢٥

وذلك باستخدام :

١ - القيم الاصلية ٢ - الفروق البسيطة ٣ - الفروق المعدلة

الحل :

١ – لحساب الانحراف المعيارى باستخدام القيم الاصلية نكون الجدول التالى :

$س_ر$	$س_ر^٢$
٦٢٥	٣٩٠٦٢٥
٧٠٠	٤٩٠٠٠٠
٦٥٠	٤٢٢٥٠٠
٧٢٥	٥٢٥٦٢٥
٢٧٠٠	١٨٢٨٧٥٠

$$\therefore ع = \sqrt{\frac{مجـ س_ر^٢}{ن} - \left(\frac{مجـ س}{ن}\right)^٢}$$

$$= \sqrt{\frac{١٨٢٨٧٥}{٤} - \left(\frac{٢٧٠٠}{٤}\right)^٢} = ٣٩,٥٣$$

٢ – لحساب الانحراف المعيارى باستخدام طريقة الفروق البسيطة نكون الجدول التالى وذلك بعد أخذ القيمة (٦٥٠) كوسط فرضى .

$س_ر$	ح = ($س_ر$ - ٦٥٠)	$ح^٢$
٦٢٥	-٢٥	٦٢٥
٧٠٠	٥٠	٢٥٠٠
٦٥٠	صفر	صفر
٧٢٥	٧٥	٥٦٢٥
-	١٠٠	٨٧٥٠

$$\therefore ع = \sqrt{\frac{مجـ\ ح^٢}{ن} - \left(\frac{مجـ\ ح}{ن}\right)^٢}$$

$$= \sqrt{\frac{٨٧٥٠}{٤} - \left(\frac{١٠٠}{٤}\right)^٢} = ٣٩,٥٣$$

٣ - أخيرا ولحساب الانحراف المعيارى باستخدام طريقة الفـروق المعدلـة

نكون الجدول التالى بعد اخذ قيمة أ = ٦٥٠ ، ب = ٢٥

س	ح	$ح' = \frac{ح}{٢٥}$	$ح'^{٢}$
٦٢٥	٢٥ -	١-	١
٧٠٠	٥٠	٢	٤
٦٥٠	صفر	صفر	صفر
٧٢٥	٧٥	٣	٩
-	-	٤	١٤

$$\therefore ع = ب \sqrt{\frac{مجـ ح'^{٢}}{ن} - \left(\frac{مجـ ح'}{ن}\right)^{٢}}$$

$$= ٢٥ \sqrt{\frac{١٤}{٤} - \left(\frac{٤}{٤}\right)^{٢}} = ٢٥ \sqrt{٢,٥} = ٣٩,٥٣$$

* لاحظ أن :-

الانحراف المعيارى لـه هـو نفـس النـاتج المتحصـل عليـه مـن الطـرق الثلاثة السابقة .

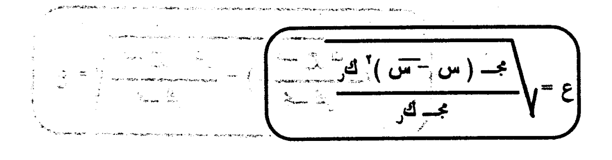
ثانيا : الانحراف المعيارى فى حالة البيانات المبوبة

(التوزيعات التكرارية)

يمكن حساب الانحراف الميعارى من بيانات مبوبة فى صورة توزيع تكرارى بأى من الطرق الثلاثة السابقة وهى الطريقة العادية أو طريقة الفروق البسيطة أو طريقة الفروق المعدلة مع الأخذ فى الاعتبار أن قيم سر تمثل مراكز الفئات كما هو الحال عند حساب الوسط الحسابى

*** الطريقة العادية :-**

تقوم هذه الطريقة على اساس ايجاد انحراف مركز الفئات (سر)عن وسطها الحسابى ($\bar{س}$) ونربعه ثم نضرب كل فرق فى التكرار المناظر لكل فئة ونوجد المجموع ويقسمة الناتج على مجموع التكرارات (مجـ كر) ثم أخذ الجذر التربيعى للمقدار الناتج نحصل على الانحراف المعيارى (ع) على اساس العلاقة التالية :



$$ع = \sqrt{\frac{مجـ (س - \bar{س})^2 كر}{مجـ كر}}$$

كما يمكن تحويل هذه الصيغة الى الصيغة التالية باسلوب مماثل لاسلوب تحويلها فى حالة البيانات الغير مبوبة :

$$\text{ع} = \sqrt{\frac{\text{مجـ س}_{\text{ر}}^{2}\ \text{ك}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}} - \left(\frac{\text{مجـ س}_{\text{ر}}\ \text{ك}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}}\right)^{2}}$$

وعادة ما يفضل استخدام الصيغة الاخيرة فى حساب الانحراف المعيارى (ع) وخاصة إذا احتوى الوسط الحسابى للتوزيع التكرارى على كسور .

<u>* طريقة الفروق البسيطة :-</u>

وهذه الطريقة تشبه الى حد كبيرا طريقة الفروق البسيطة فى حالة البيانات الغير مبوبة والتى سبق شرحها مع الأخذ فى الاعتبار التكرارات المناظرة لكل فئة .

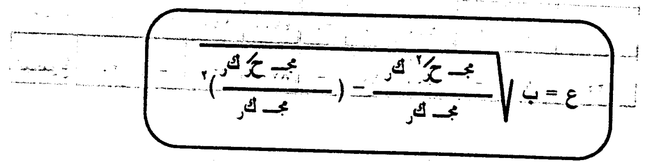
$$\text{ع} = \sqrt{\frac{\text{مجـ ح}^{2}\ \text{ك}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}} - \left(\frac{\text{مجـ ح ك}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}}\right)^{2}}$$

حيث : $\text{ح} = (\text{س}_{\text{ر}} - \text{أ})$ ، ر = ١ ، ٢ ، ... ، ن

وعادة ما يتم استخدام الصيغة السابقة فى حساب الانحراف المعيارى (ع) فى حالة التوزيعات التكرارية المنتظمة والغير منتظمة .

*** طريقة الفروق المعدلة :-**

يفضل استخدام هذه الطريقة فى حساب الانحراف المعيارى فى الحالات التى تقبل فيها جميع قيم الفروق البسيطة القسمة على مقدار ثابت (ب) بدون باقى وحيث أن (ع) يتأثر بعملية الضرب فإن :



مثال (٩) :-

باستخدام كل طريقة من الطرق الثلاثة التالية :

١ – الطريقة العادية ٢– طريقة الفروق البسيطة

٣ – طريقة الفروق المعدلة

احسب الانحراف المعيارى للتوزيع التكرارى التالى :-

فئات	صفر –	١٠ –	٢٠ –	٣٠–٤٠
تكرارات	٥	٨	٣	٤

الحل

لايجاد الانحراف المعيارى بالطرق الثلاثة السابقة نكون الجدول التالى :

فئات	تكرارات	الطريقة العادية			طريقة الفروق البسيطة			طريقة الفروق المعدلة		
	($ك_ر$)	$س_ر$	$س_ر ك_ر$	$س_ر^2 ك_ر$	$ح$	$ح ك_ر$	$ح^2 ك_ر$	$ح'$	$ح' ك_ر$	$ح'^2 ك_ر$
صفر-	٥	٥	٢٥	١٢٥	-١٠	-٥٠	٥٠٠	-١	-٥	٥
١٠ -	٨	١٥	١٢٠	١٨٠٠	صفر	صفر	صفر	صفر	صفر	صفر
٢٠ -	٣	٢٥	٧٥	١٨٧٥	١٠	٣٠	٣٠٠	١	٣	٣
٣٠-٤٠	٤	٣٥	١٤٠	٤٩٠٠	٢٠	٨٠	١٦٠٠	٢	٨	١٦
المجموع	٢٠	-	٣٦٠	٨٧٠٠	-	٦٠	٢٤٠٠	-	٦	٢٤

١- الطريقة العادية :

يحسب الانحراف المعيارى من الجزء الاول الخاص بهذه الطريقة من الجدول السابق وبالتعويض فى المعادلة التالية ينتج أن :

$$ع = \sqrt{\frac{\text{مجـ } س_ر^2 ك_ر}{\text{مجـ } ك_ر} - \left(\frac{\text{مجـ } س_ر ك_ر}{\text{مجـ } ك_ر}\right)^2}$$

$$= \sqrt{\frac{٨٧٠٠}{٢٠} - (\frac{٣٦٠}{٢٠})^{٢}} = ١٠,٥٣٩$$

٢- طريقة الفروق المبسطة :

وفيها يحسب الانحراف المعيارى من الجزء الثانى الخاص بهذه الطريقة من الجدول السابق وبالتعويض فى المعادلة التالية ينتج أن :

$$ع = \sqrt{\frac{مجـ\ ح^{٢}\ ك ر}{مجـ\ ك ر} - (\frac{مجـ\ ح\ ك ر}{مجـ\ ك ر})^{٢}}$$

$$= \sqrt{\frac{٢٤٠٠}{٢٠} - (\frac{٦٠}{٢٠})^{٢}} = ١٠,٥٣١$$

٣- طريقة الفروق المعدلة :

وفيها يحسب الانحراف المعيارى من الجزء الثالث الخاص بهذه الطريقة من الجدول السابق وبالتعويض فى المعادلة التالية ينتج أن :

$$ع = ب \sqrt{\frac{مجـ\ ح^{٢}\ ك ر}{مجـ\ ك ر} - (\frac{مجـ\ ح\ ك ر}{مجـ\ ك ر})^{٢}}$$

$$= ١٠ \sqrt{\frac{٢٤}{٢٠} - (\frac{٦}{٢٠})^{٢}} = ١٠,٥٣٦$$

لوحظ وجود اختلاف بسيط فى قيم الانحراف المعيارى باستخدام كل طريقة ويرجع سبب ذلك الى اختلاف الاساس الرياضى الذى تقوم عليه كل طريقة من الطرق الثلاثة السابقة .

كما يتضح مما سبق عرضه أن الانحراف المعيارى إكثر ملائمة من الانحراف المتوسط لأسباب كثيره منها استجابته الكبيرة للعمليات الحسابية بعكس الانحراف المتوسط ، ولكن ما زالت جميع مقاييس التشتت المطلقة (المدى – الانحراف الربيعى – الانحراف المتوسط – الانحراف المعيارى) تشترك جميعها فى أن لها تمييز ولا تصلح للمقارنة بين المجموعات اذا كانت وحدات القياس الخاصة بهذه المجموعات مختلفة .

ومما ينبغى ملاحظته أنه يمكننا حساب الانحراف المعيارى للتوزيعات التكرارية النسبية بعد تعديل الصيغ الثلاثة السابقة للتوزيعات التكرارية على النحو التالى :

١ – معادلة حساب الانحراف المعيارى (ع) __بالطريقة العادية__ يمكن تعديلها الى الصيغة التالية :

$$ع = \sqrt{مجـ\ س_ر^{٢}\ د_ر - (مجـ\ س_ر\ د_ر)^{٢}}$$

حيث س : تمثل مراكز الفئات ، د : تمثل التكرار النسبى = $\frac{\text{ك}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}}$

٢ – معادلة حساب الانحراف الميعارى (ع) وفقا لطريقة الفروق البسيطة يمكن تعديلها الى الصيغة التالية :

$$\text{ع} = \sqrt{\text{مجـ ح}^{٢}\,\text{د}_{\text{ر}} - (\text{مجـ ح د}_{\text{ر}})^{٢}}$$

حيث : ح = ($\text{س}_{\text{ر}}$ – أ)   ، أ عبارة عن الوسط الفرضى

٣ – معادلة حساب الانحراف الميعارى (ع) وفقا لطريقة الفروق المعدلة يمكن تعديلها الى الصيغة التالية :

$$\text{ع} = \text{ب}\sqrt{\text{مجـ ح}'^{٢}\,\text{د}_{\text{ر}} - (\text{مجـ ح}'\,\text{د}_{\text{ر}})^{٢}}$$

حيث : $\text{ح}' = \frac{\text{ح}}{\text{ب}}$

مثال (١٠) :

احسب الانحراف المعيارى (ع) بالطرق الثلاثة (العادية والفروق البسيطة والفروق المعدلة) للتوزيع النسبى التالى لدرجات عشرين طالب فى الرياضيات :

فئات الدرجات	صفر -	١٠ -	٢٠ -	٣٠ - ٤٠
التكرار النسبى	,٢٥	,٤٠	,١٥	,٢٠

الحل :

لايجاد قيمة الانحراف المعيارى (ع) للتوزيع النسبى لدرجات عشرين طالب فى مادة الرياضيات بالطرق الثلاثة السابقة نكون الجدول التالى :

فئات الدرجات	تكرار نسبى (در)	الطريقة العادية			طريقة الفروق البسيطة			طريقة الفروق المعدلة		
		س	س در	$س^2$ در	ح	ح در	$ح^2$ در	ح/	ح/ در	$ح/^2$ در
صفر	٠,٢٥	٥	١,٢٥	٦,٢٥	١٠-	٢,٥-	٢٥	١-	٠,٢٥-	,٢٥
١٠ -	٠,٤٠	١٥	٦,٠٠	٩٠,٠٠	صفر	صفر	صفر	صفر	صفر	صفر
٢٠ -	٠,١٥	٢٥	٣,٧٥	٩٣,٧٥	١٠	١,٥	١٥	١	٠,١٥	٠,١٥
٣٠ - ٤٠	٠,٢٠	٣٥	٧,٠٠	٢٤٥,٠٠	٢٠	٤,٠٠	٨٠	٢	٠,٤٠	٠,٨٠
المجموع	١,٠٠	-	١٨,٠٠	٤٣٥,٠٠	-	٣,٠٠	١٢٠	-	٠,٣	١,٢

١- الطريقة العادية :

يتم حساب الانحراف المعيارى (ع) من الجزء الاول الخاص بهذه الطريقة من الجدول السابق وبالتعويض فى المعادلة التالية ينتج أن :

$$ع = \sqrt{\text{مجـ } س_{ر}^{٢} \, د_{ر} - (\text{مجـ } س_{ر} \, د_{ر})^{٢}}$$

$$= \sqrt{٤٣٥ - (١٨)^{٢}} = \sqrt{١١١} = ١٠,٥٣٦$$

٢- طريقة الفروق البسيطة :

وفيها يحسب الانحراف المعيارى من الجزء الثانى الخاص بهذه الطريقة من الجدول السابق وبالتعويض فى المعادلة التالية ينتج أن :

$$ع = \sqrt{\text{مجـ } ح_{ر}^{٢} \, د_{ر} - (\text{مجـ } ح_{ر} \, د_{ر})^{٢}}$$

$$= \sqrt{١٢٠ - (٣)^{٢}} = \sqrt{١١١} = ١٠,٥٣٦$$

٣- طريقة الفروق المعدلة :

وفيها يحسب الانحراف المعيارى من الجزء الثالث الخاص بهذه الطريقة من الجدول السابق وبالتعويض فى المعادلة التالية ينتج أن :

$$ع = \sqrt{\text{مجـ } ح'^{٢}_{ر} \, د_{ر} - (\text{مجـ } ح'_{ر} \, د_{ر})^{٢}}$$

$$= ١٠ \sqrt{١,٢ - (,٣)^{٢}} = ١٠,٥٣٦$$

*** ملاحظات عامة على التباين والانحراف المعيارى :-**

يعد التباين وبالتالى الانحراف المعيارى من المقاييس التى تأخذ فى الاعتبار جميع المفردات عند قياس التشتت أو الانتشار ، علاوة على انهما يعطيان وزنا أكبر للمفردات المتطرفة . هذا فى حين نجد أن نصف المدى الربيعى أقل هذه المقاييس تأثرا بالقيم المتطرفة وهو لهذا السبب يفضل عن بقية المقاييس عند قياس تشتت توزيع شديد الالتواء .

- كذلك يعد الانحراف المعيارى من أكثر مقاييس التشتت استخداما فى الاحصاء وأكثرها أهمية وذلك لاعتبارات متعددة تتعلق بعملية الاستنتاج الاحصائى سواء فى نظرية التقديرات أو إختبارات الفروض .
- يدخل الانحراف المعيارى فى تركيب العديد من المقاييس الاحصائية الهامة مثل مقاييس الالتواء والاختلاف النسبى والدرجات المعيارية ومعامل الارتباط كما سنبين فى الاجزاء التالية من هذا المؤلف.
- أخيرا تم استنتاج علاقات تجريبية تربط بين الانحراف المعيارى ونصف المدى الربيعى والانحراف المتوسط ويشترط لتحقيق هذه العلاقات أن يكون التوزيع متماثل أو بسيط (محدود) الالتواء وتقضى هذه العلاقات بأن :

الانحراف المتوسط = $\frac{4}{5}$ الانحراف المعيارى تقريبا

نصف المدى الربيعى = $\frac{2}{3}$ الانحراف المعيارى تقريبا

ويمكن الاعتماد على هذه العلاقة ، إما للتحقق السريع من صحة حسابنا لنصف المدى الربيعى أو لتقدير قيمة الانحراف الميعارى وكذلك التباين خاصة فى حالة التوزيعات التكرارية المفتوحة ، حين يتعذر حسابهما منها .

(٤-١-٤-١) التباين المصحح لشبرد :-

يعتمد كثيرا من المقاييس الاحصائية مثل الوسط الحسابى والتباين فى حسابهما على مراكز الفئات وأيضا طول الفئة ولذلك فعند حساب تلك المقاييس من بيانات أصلية وبيانات مبوبة نجد أن هناك فروقا خاصة عند حساب التباين ويزداد هذا الفرق بزيادة طول الفئة . ولذا يستلزم الامر تصحيح التباين الناتج من الجداول التكرارية ، ولقد قام شبرد (*Sheppard*) بعمل هذا التصحيح كالاتى :

$$ع'^{2} = \text{التباين المصحح لشبرد} = ع^{2} - \frac{ب^{2}}{12}$$

وبالتالى فإن الانحراف المعيارى المصحح للشبرد هو :

$$ع' = \sqrt{ع^{2} - \frac{ب^{2}}{12}}$$

حيث $ع^{2}$ = التباين المبيانات المبوبة ، ب = طول الفئة

مثال (١١) :-

فى المثال السابق أوجد الانحراف المعيارى المصحح لدرجات الطلاب فى مادة الرياضيات .

الحل :

$$ع^- = \sqrt{(١٠,٥٣٦)^٢ - \frac{١٠٠}{١٢}} = \sqrt{١٠٢,٦٧٣} = ١٠,١٣٢$$

ومما ينبغى ملاحظته وجود تصحيحات مماثلة لتصحيح شبرد بالنسبة للتوزيعات الاخرى لا يتسع المجال الآن لعرضها .

(٤-٣) مقاييس الاختلاف النسبى والقيم المعيارية:-

(٤-٣-١) مقاييس الاختلاف النسبى *Measures of Relative Variation*

قد نرغب فى بعض الاحيان فى مقارنة مجموعات مختلفة من البيانات لا من حيث درجة تجانسها المطلقة فحسب ولكن من حيث درجة تجانسها النسبى خاصة حين تختلف المتوسطات الحسابية لها . إذ أنه من المتوقع أن تكون المجموعات ذات المتوسطات الحسابية الكبيرة ذات انحرافات معيارية كبيرة نسبيا لذلك فإن الاقتصار على استخدام الانحراف المعيارى فى المقارنة فى حالات كهذه قد يقودنا الى استنتاجات خاطئة .

فإذا أخذنا على سبيل المثال فصلين دراسيين مختلفين وكان متوسط الدرجات فى مادة الاحصاء فى شهر معين فى الفصل الاول هو ٢٦ درجة فى حين أن متوسط الدرجات لنفس المادة فى الفصل الدراسى الثانى هو ٣٨ درجة وإذا كان الانحراف المعيارى لتوزيع الدرجات فى الفصل الاول هو ٣درجة وفى الفصل الثانى هو ٥ درجات ، فإذا إقتصرت المقارنة على درجة الاختلاف للفصل لكانت النتيجة التى تصل أليها أن الاختلاف المطلق فى الفصل الثانى الى الاول كنسبة ٥:٣ ولكننا إذا استخدمنا فى المقارنة الحجم النسبى للاختلاف او الانحراف المعيارى منسوبا الى الوسط الحسابى لكان الاختلاف النسبى لتوزيع الدرجات فى الفصل الثانى الى الاول كنسبة ١١,٥٪ الى ١٣,٢٪ والفرق بين النتيجتين واضح . والاسلوب الذى اتبع هنا هو نسبة مقياس التشتت (أو الاختلاف المطلق) الى مقياس مناظر للقيمة المتوسطة – وهذه النسبة وهى مقياس للأختلاف النسبى – تعرف باسم معامل الاختلاف .

وهناك عده صور لمقاييس التشتت النسبية (معاملات الاختلاف) تعتمد كلها على نسبة مقياس من مقاييس التشتت الى مقياس من مقاييس النزعة المركزية وضرب الناتج ×١٠٠ نذكر منها :

١ – معامل الاختلاف = $\frac{ع}{\overline{س}} \times ١٠٠$

$$\text{٢ – معامل الاختلاف} = \frac{\text{ر٣} - \text{ر١}}{\text{ر٢}} \times ١٠٠$$

أو

$$\text{معامل الاختلاف} = \frac{\text{ر٣} - \text{ر١}}{\text{ر٣} + \text{ر١}} \times ١٠٠$$

وعادة ما يتم استخدام احدى الصورتين السابقتين فى حالة الجداول التكرارية المفتوحة .

$$\text{٣ – معامل الاختلاف} = \frac{\text{الانحراف المتوسط عن الوسط الحسابى (أو الوسيط)}}{\text{الوسط الحسابى (أو الوسيط)}} \times ١٠٠$$

ويجب التنوية إلى أنه يمكننا استخدام أى من المقاييس السابقة لدراسة الاختلاف النسبى لتوزيع ظاهرة ما فى عدة مجتمعات أو عدة ظواهر فى مجتمع واحد خاصة حين تختلف وحدات القياس الخاصة بتلك الظواهر ، مما يجعل المقارنة بين اختلافاتها المطلقة غير منطقى .

وغنى عن البيان أنه لمقارنة توزيعين مختلفين من حيث درجة الاختلاف النسبى فأنه يجب استخدام معامل اختلاف واحد فقط للمقارنة مالم توجد مقاييس مشتركة للمقارنة بين المجموعات .

مثال (١٢) :-

قارن بين تشتت المجموعتين أ ، ب اذا علمت أن :

المقياس / المجموعة	أ	ب
الوسط الحسابي	٢٠	٤٠
الانحراف المعيارى	٢	٣

الحل :

معامل الاختلاف للمجموعة (أ) = $\frac{ع}{\bar{س}}$ × ١٠٠

= $\frac{٢}{٢٠}$ × ١٠٠ = ١٠٪

معامل الاختلاف للمجموعة (ب) = $\frac{ع}{\bar{س}}$ × ١٠٠

= $\frac{٣}{٤٠}$ × ١٠٠ = ٧,٥٪

مما ينبغى ملاحظتة أن هذا المثال قد مكننا من التعرف على الخطأ الذى يمكن أن نقع فيه إذا ما إعتمدنا فى عملية المقارنة بين تشتت المجموعتين على اساس الانحراف المعيارى فقط كمقياس للتشت المطلق حيث أوضحت المقارنة بين تشتتى المجموعتين بناء على معامل الاختلاف أن تشتت المجموعة الاولى اقل من

تشتت المجموعة الثانية وهذا مخالف لما أوضحته نتيجة المقارنة بين تشتتى المجموعتين على اساس الانحراف المعيارى فقط .

مثال(١٣) :-

البيانات التالية توضح الاجور اليومية لعينة مكونة من ١٠٠ عامل فى إحدى المؤسسات الصناعية :

الاجر اليومي بالجنية	أقل من ٣٥	٣٥ -	٣٨ -	٤١ -	أكبر من ٤٣
عدد العمال	١٤	٢٧	٣٠	٢٠	٩

المطلوب :

حساب مقياس مناسب للتشتت لمقارنته بمقياس أخر لدراسة تشتت عينة أخرى وكانت بياناتها المتاحة على النحو التالى :

$ر_١$ = ٢٥ ، $ر_٣$ = ٤٥ علما بأن الحد الأدنى للأجور هو ٣٠ جنيه فى اليوم .

الحل :

فئات الأجر	عدد العمال	حدود عليا للفئات	تكرار متجمع صاعد	ملاحظات
٣٠-٣٥	١٤	أقل من ٣٥	١٤	
				موقع ترتيب $ر_1$
٣٥ -	٢٧	أقل من ٣٨	٤١	
٣٨ -	٣٠	أقل من ٤١	٧١	
				موقع ترتيب $ر_3$
٤١ -	٢٠	أقل من ٤٣	٩١	
أكبر من ٤٣	٩	٤٣ فأكثر	١٠٠	
المجموع	١٠٠	—	—	—

نظرا لأن التوزيع التكرارى السابق مفتوح وبالتالى يفضل استخدام الصوره التالية لمعامل الاختلاف والتى تتفق أيضا والبيانات المتاحة عن العينة الثانية .

$$\text{معامل الاختلاف} = \frac{ر_3 - ر_1}{ر_3 + ر_1} \times ١٠٠$$

العينة الاولى :

أولا ايجاد قيم $ر_1$ ، $ر_3$ للظاهرة الاولى :

$$\text{ترتيب } ر_1 = \frac{\text{مجـ ك}}{٤} = ٢٥ \quad \text{فئة } ر_1 \text{ هى } (٣٥ - ٣٨)$$

$$\text{ر}_1 = \text{ف د}_1 + \frac{\frac{\text{مجـ ك}}{٤} - \text{ك د}_1}{\text{ك أ}_1 - \text{ك د}_1} \times \text{ل}_1$$

$$= ٣٥ + \frac{٢٥ - ١٤}{٤١ - ١٤} \times ٣ = ٣٧,٠٤$$

$$\text{ترتيب ر}_3 = \frac{٣ \text{ مجـ ك}}{٤} = \frac{٣ \times ١٠٠}{٤} = ٧٥$$

فئة ر_3 هى (٤١ - ٤٣)

$$\text{ر}_3 = \text{ف د}_3 + \frac{\frac{٣ \text{ مجـ ك}}{٤} - \text{ك د}_3}{\text{ك أ}_3 - \text{ك د}_3} \times \text{ل}_3$$

$$= ٤١ + \frac{٧٥ - ٧١}{٩١ - ٧١} \times ٢ = ٤١,٤$$

$$\text{معامل الاختلاف للعينة الاولى} = \frac{٤١,٤ - ٣٧,٠٤}{٤١,٤ + ٣٧,٠٤} \times ١٠٠ = ٥,٥٦\%$$

العينة الثانية :

$\text{ر}_1 = ٢٥$ ، $\text{ر}_3 = ٤٥$

$$\text{معامل الاختلاف للعينة الثانية} = \frac{٤٥ - ٢٥}{٤٥ + ٢٥} \times ١٠٠ = ٢٨,٥٧\%$$

أى أن الاجور فى العينة الاولى أكثر تجانسا (أقل تشتت) من الأجور فى العينة الثانية .

مثال(١٤) :-

فيما يلى التوزيعين التكراريين للأجور الشهرية وكميات الانتاج الاسبوعية لمجموعتين من العمال :

المجموعة الاولى:

فئات الأجور	٥ -	٧ -	٩ -	١١ -	١٣ - ١٥	المجموع
عدد العمال	٤	٨	٢٧	٦	٥	٥٠

المجموعة الثانية :

فئات الأنتاج	٢٠ -	٣٠ -	٤٠ -	٥٠ -	٦٠ - ٧٠	المجموع
عدد العمال	٦	١٠	٢٢	٧	٥	٥٠

المطلوب : مقارنة تشتتى مجموعتى الاجور والانتاج .

ارشادات الحل :

حيث لم نذكر الصورة المطلوبة لمعامل الاختلاف صراحة ؛ فإن الصورة المستخدمة فى المقارنة هى :

$$\text{معامل الاختلاف} = \frac{ع}{\overline{س}} \times ١٠٠$$

وتكون نتيجة المقارنة كما يلى :

$$\text{معامل الاختلاف لمجموعة الأجور} = \frac{ع}{\overline{س}} \times ١٠٠ = ٢٠\%$$

$$\text{أما معامل الاختلاف لمجموعة الانتاج} = \frac{ع}{\overline{س}} \times ١٠٠ = ٣٥\%$$

مما يعنى أن تشتت مجموعة الانتاج أكبر من تشتت مجموعة الاجور وذلك بناء على قيم معامل الاختلاف للمجموعتين .

(٤-٣-٣) القيم المعيارية *Standardized Values* :-

أوضحنا فيما سبق كيف يمكن المقارنــة بين تشتت مجموعتين أو أكثر ، مجموعتين إلا أننا قد نحتاج فى بعض الأحيان الى مقارنة مفردتين تنتميان الى مجموعتين مختلفتين . فهذه الحالة تبرز الحاجة للوصول الى مقياس احصائى يأخذ فى الاعتبار الترتيب النسبى أو المركز النسبى لكل من المفردتين داخل المجموعة التى تنتمى اليها .

مثال (١٥) :-

اذا حصل طالب على ٨٠ درجة فى مادة الاحصاء ، ٧٠ درجة فى مادة الاقتصاد واقتصرنا فى المقارنة على الدرجتين فقط دون اعتبار لتوزيع درجات كل من هاتين المادتين لقلنا أن أداء الطالب فى مادة الأحصاء أفضل منه فى مادة الاقتصاد .وإذا علمنا أن الوسط الحسابى لدرجات الطلاب فى مادة الاحصاء هو $\bar{س} = ٧٠$ درجة ولمادة الاقتصاد هو $\bar{س} = ٦٥$ درجة وكانت الانحرافات المعيارية هى $ع_1 = ٥$ درجات ، $ع_2 = ٢$ درجة فى الأحصاء والأقتصاد على التوالى .

فإذا أخذنا المركز النسبى لدرجة الطالب فى كل ماده بالنسبة لتوزيع درجات هذه الماده على اساس الدرجة المعيارية وهى :

$$ص = \frac{س - \bar{س}}{ع}$$

يتبين لنا أن الدرجة المعيارية ($ص_1$) والتى تناظر درجة النجاح فى مادة الاحصاء (٨٠ درجة) هى :

$$ص_1 = \frac{٨٠ - ٧٠}{٥} = ٢ \text{ درجة معيارية}$$

بينما الدرجة المعيارية ($ص_2$) والتى تناظر درجة النجاح فى مادة الاقتصاد (٧٠ درجة) هى :

$$ص_2 = \frac{٧٠ - ٦٥}{٢} = ٢,٥ \text{ درجة معيارية}$$

من المثال (١٥) يمكننا استنتاج أن أداء الطالب كان أفضل فى مادة الاقتصاد عنه فى مادة الاحصاء وهو عكس الاستنتاج السابق تماما.

ومن المعروف أن القيم المعيارية لها مجموعة خصائص منها :

- تنحصر هذه القيم بين -٣ ، +٣ فى جميع المجموعات .
- الوسط الحسابى لها دائما يساوى صفر بغض النظر عن الوسط الحسابى الأصلى لمجموعة البيانات الأصلية .
- الانحراف المعيارى للقيم المعيارية يساوى دائما الواحد الصحيح .

مثال عام:

الجدول التالى يوضح التوزيع التكرارى لأوزان ٨٠ شخصا

الوزن بالكيلو جرام	٦٢ -	٦٦ -	٧٠ -	٧٤ -	٧٨ -	٨٢ -	٨٦ - ٩٠
عدد الأشخاص	٣	٨	٢٠	٢١	١٤	١٠	٤

المطلوب :

١ - حساب قيمة الوسط الحسابى والانحراف المعيارى للتوزيع .

٢ - حساب قيمة الوسيط والانحراف الربيعى للتوزيع .

٣ - حساب قيمة المنوال للتوزيع .

٤ - حساب معامل الاختلاف .

٥ - هل هذا التوزيع متماثل ؟ علل اجابتك .

الحل

1 - حساب قيمة الوسط الحسابى ($\bar{س}$) والانحراف المعيارى (ع)

فئات الوزن بالكيلو جرام	عدد الأشخاص (التكرار = $ك_ر$)	مراكز الفئات ($س_ر$)	الفروق البسيطة $ح = (س_ر - أ)$	الفروق المعدلة $ح' = \frac{ح}{ب}$	$ح' ك_ر$	$ح'^2 ك_ر$
٦٢ -	٣	٦٤	١٢ -	٣ -	٩ -	٢٧
٦٦ -	٨	٦٨	٨ -	٢ -	١٦ -	٣٢
٧٠ -	٢٠	٧٢	٤ -	١-	٢٠ -	٢٠
٧٤ -	٢١	٧٦	صفر	صفر	صفر	صفر
٧٨ -	١٤	٨٠	٤	١	١٤	١٤
٨٢ -	١٠	٨٤	٨	٢	٢٠	٤٠
٨٦ - ٩٠	٤	٨٨	١٢	٣	١٢	٣٦
المجموع	٩٠	-	-	-	٤٥ - ٤٦+ ١ +	١٦٩

$$\text{الوسط الحسابى } (\bar{س}) = أ + \frac{مجـ\ ح' ك_ر}{مجـ\ ك_ر} \times ب$$

$$= \text{٧٦} + \frac{\text{١}}{\text{٨٠}} \times \text{٤} = \text{٧٦,٠٥} \ \text{كيلو جرام}$$

الانحراف المعيارى (ع)

$$\text{ع} = \text{ب} \sqrt{\frac{\text{مجـ } \text{ح}'^{\text{٢}} \text{ ك}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}} - \left(\frac{\text{مجـ ح}' \text{ ك}_{\text{ر}}}{\text{مجـ ك}_{\text{ر}}}\right)^{\text{٢}}}$$

$$= \text{٤} \sqrt{\frac{\text{١٦٩}}{\text{٨٠}} - \left(\frac{\text{١}}{\text{٨٠}}\right)^{\text{٢}}}$$

$$= \text{٤} \sqrt{\text{٢,٠٩٩٩}} = \text{٤} \times \text{١,٤٤٩١} = \text{٥,٨} \ \text{كيلو جرام}$$

٢ - حساب قيمة الوسيط ($\text{ر}_{\text{٢}}$) والانحراف الربيعى (نصف المدى الربيعى) للتوزيع

، نكون جدول تكرارى متجمع صاعد ومنه نحسب المطلوب كما يلى :

جدول تكرارى متجمع صاعد لأوزان ٨٠ شخصا

فئات الوزن	عدد الأشخاص (التكرار = $\text{ك}_{\text{ر}}$)	حدود عليا للفئات	تكرار متجمع صاعد	ملاحظات
٦٢ -	٣	أقل من ٦٦	٣	
٦٦ -	٨	أقل من ٧٠	١١	
			←	موقع ترتيب $\text{ر}_{\text{١}}$
٧٠ -	٢٠	أقل من ٧٤	٣١	
			←	موقع ترتيب $\text{ر}_{\text{٢}}$
٧٤ -	٢١	أقل من ٧٨	٥٢	
			←	موقع ترتيب $\text{ر}_{\text{٣}}$
٧٨ -	١٤	أقل من ٨٢	٦٦	
٨٢ -	١٠	أقل من ٨٦	٧٦	
٨٦ - ٩٠	٤	أقل من ٩٠	٨٠	
المجموع	٨٠	—	—	

(ب) حساب قيمة الوسيط ($ر_٢$) :

ترتيب ($ر_٢$) = $\frac{مجـ ك_ر}{٢} = \frac{٨٠}{٢} = ٤٠$ ، فئة الوسيط هى (٧٤ - ٧٨)

$$\therefore \text{قيمة } ر_٢ = ف\, ر_٢ + \frac{\frac{مجـ ك}{٢} - ك\, د_٢}{ك\, أ_٢ - ك\, د_٢} \times ل_٢$$

$$= ٧٤ + \frac{٤٠ - ٣١}{٥٢ - ٣١} \times ٤ = ٧٥,٧ \text{ كيلو جرام}$$

(حـ) حساب قيمة الانحراف الربيعى (نصف المدى الربيعى) ، لابد من حساب قيمة كل من الربيع الثالث ($ر_٣$) والربيع الاول ($ر_١$) أولا ، ثم التعويض فى الصيغة الخاصة بنصف المدى الربيعى فنحصل على المطلوب كما يلى :

١- حساب قيمة الربيع الاول ($ر_١$)

ترتيب $ر_١$ = $\frac{مجـ ك}{٤} = \frac{٨٠}{٤} = ٢٠$ ، فئة $ر_١$ هى (٧٠ - ٧٤)

$$\therefore \text{قيمة } ر_١ = ف\, د_١ + \frac{\frac{مجـ ك}{٤} - ك\, د_١}{ك\, أ_١ - ك\, د_١} \times ل_١$$

$$= ٧٠ + \frac{٢٠ - ١١}{٣١ - ١١} \times ٤ = ٧١,٨ \text{ كيلو جرام}$$

٢ - حساب قيمة الربيع الثالث ($ر_٣$)

$$\text{ترتيب } ر_٣ = \frac{٣ \text{ مجـ ك}}{٤} = \frac{٣ \times ٨٠}{٤} = ٦٠ \text{ ، فئة } ر_٣ \text{ هى } (٧٨-٨٢)$$

$$\therefore \text{ قيمة } ر_٣ = \text{ف د}_٣ + \frac{\frac{٣ \text{ مجـ ك}}{٤} - \text{ك د}_٣}{\text{ك أ}_٣ - \text{ك د}_٣} \times ل_٣$$

$$= ٧٨ + \frac{٦٠ - ٥٢}{٦٦ - ٥٢} \times ٤ = ٨٠,٣ \text{ كيلوجرام}$$

$$\therefore \text{ الانحراف الربيعى (نصف المدى الربيعى)} = \frac{ر_٣ - ر_١}{٢}$$

$$= \frac{٨٠,٣ - ٧١,٨}{٢} = ٤,٣ \text{ كيلو جرام}$$

٣ - حساب قيمة المنوال للتوزيع (بطريقة الفروق) بيرسون .

نظرا لتساوى أطوال فئات التوزيع (منتظم) ، إذن لابد من استخدام التكرارات الاصلية فى ايجاد قيمة المنوال على النحو التالى :

فئة المنوال هى (٧٤ - ٧٨) ⟵⟶ وهى المناظرة لاكبر تكرار

قيمة المنوال = بداية الفئة المنوالية + س

م = ٧٤ + س

حيث س يتم حسابها على النحو التالى :

$$\frac{\text{س}}{(\text{طول الفئة المنوالية} - \text{س})} = \frac{\text{تكرار الفئة المنوالية} - \text{التكرار السابق لها}}{\text{تكرار الفئة المنوالية} - \text{التكرار اللاحق لها}}$$

$$\frac{\text{س}}{(٤ - \text{س})} = \frac{٢٠ - ١١}{٢١ - ١٤}$$

$$\frac{\text{س}}{(٤ - \text{س})} = \frac{١}{٧}$$

$$٧ \text{ س} = ٤ - \text{س}$$

$$٨ \text{ س} = ٤$$

$$\text{س} = \frac{٤}{٨} = ,٥$$

المنوال (م) = ٧٤ + ,٥ = ٧٤,٥ كيلو جرام

٤ - حساب معامل الاختلاف :

$$\text{معامل الاختلاف} = \frac{\text{ع}}{\overline{\text{س}}} \times ١٠٠$$

$$= \frac{٥,٨}{٧٦,٠} \times ١٠٠ = ٧,٦\%$$

٥ - لمعرفة ما إذا كان التوزيع (الاوزان) متماثل أم لا ، نوجد الالتواء للتوزيع ،

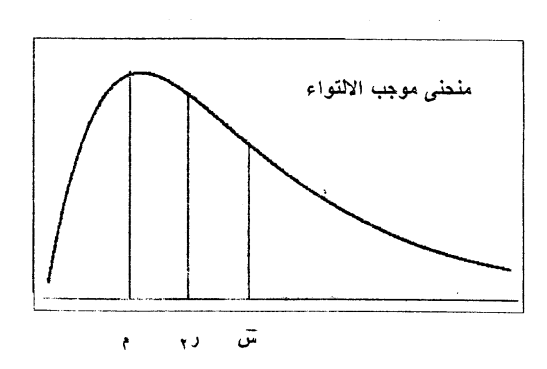
فإذا كان الناتج يساوى صفر دل ذلك على أن التوزيع متماثل ، وإذا كان غير

ذلك كان التوزيع غير متماثل (ملتو موجب أو ملتو سالب) ولتحقيق ذلك نعوض فى العلاقة التالية .

(ر٣ − ر٢) − (ر٢−ر١) = (٨٠,٣ − ٧٥,٧) − (٧٥,٧ − ٧١,٨)

= ٤,٦ − ٣,٩ = + ٠,٧

أى أن التوزيع غير متماثل لأن ناتج التعويض لا يساوى الصفر ، بل ملتو جهة اليمين لأن (أشارة معامل الالتواء موجبة) ويأخذ منحنى التوزيع الشكل التالى :



وهذا ما يتفق مع نتائج حساب كلا من الوسط الحسابى والوسيط والمنوال لتوزيع الأوزان .

الباب الخامس

العزوم والالتواء والتفرطح

Moments & Skewness & Kurtosis

درسنا فى البابين السابقين خاصيتين أساسيتين للتوزيعات التكرارية وهما خاصيتى النزعة المركزية والتشتت لهذه التوزيعات والتى تساعد فى تلخيص وصف التوزيعات التكرارية ، الا أن هذه المقاييس لا تكفى وحدها للتعرف على كل خصائص التوزيعات التكرارية والمقارنة بينها ، فقد يتساوى توزيعان تكراريان من حيث القيمة المتوسطة والانحراف المعيارى ولكنهما يختلفان من حيث التماثل أو من حيث الاعتدال . لذلك فان الامر يستلزم ايجاد بعض المقاييس الاحصائية الاخرى والتى تقيس مدى تماثل التوزيع وكذا درجة تفرطحه .

وقبل أن نتعرض لدراسة خاصيتى التماثل والاعتدال بشىء من التفصيل وجب علينا أولا دراسة العزوم الرياضية بأنواعها المختلفة وكذلك أوجه الاستفادة منها فى دراسة كلا من هاتين الخاصيتين وفى قياس التشتت أيضا .

(٥ - ١) العزوم *Moments* :-

بديهيا أن كلمة عزم تشتق من علـى الاستاتيكا حيث يقـاس عـزم القـوة بحاصل ضرب هذه القوة فى ذراع عزمها (الـذراع هـو بعد عمـل خـط القـوة عـن مركز العزم) ، ويكون عزوم مجموعة من القوى = مجموع حاصل ضرب كل قوة فى ذراعها .

وهذا الاشتقاق ناتج من تشبيه التكرار بالقوة وقيمته بذراع العزم، الا أن قياس العزم فى الاحصاء يختلف عن قياسه فى الاستاتيكا .

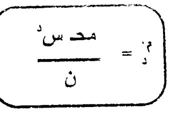
<u>* قياس العزوم الرياضية (العزوم الصفرية – العزوم المركزية) :-</u>

<u>أولاً العزوم الصفرية (حول نقطة الأصل (م))</u>

The Moments about Zero

نفرض أن س متغير يأخذ القيم س١ ، س٢ ، .... سن فى عينه حجمها (ن) . فلو رمزنا للعزم رقم دال (العزم الدالى) حول الصفر بالرمز ($م'_د$) فان :

<u>الصيغة العامة للعزوم الصفرية فى حالة البيانات الغير مبوبة</u>



$$م'_د = \frac{مجـ\ س^{د}}{ن}$$

حيث د = ١ ، ٢ ، ٣ ، ...... وبوضع د = ١ ، ٢ ، ٣ ، ٤ نحصل

على العزوم الاربعة الاولى حول الصفر على التوالى كما يلى :

$$م'_{١} = \frac{مجـ س}{ن}$$

$$م'_{٢} = \frac{مجـ س^{٢}}{ن}$$

$$م'_{٣} = \frac{مجـ س^{٣}}{ن}$$

$$م'_{٤} = \frac{مجـ س^{٤}}{ن}$$

<u>مثال (١) :-</u>

أوجد العزوم الصفرية الاربعة الاولى لمجموعة البيانات التالية :

١ ، ٣ ، ٥

الحل :

يمكن حساب العزوم الصفرية الاربعة الاولى من الجدول التالى :

س	س٢	س٣	س٤
١	١	١	١
٣	٩	٢٩	٨١
٥	٢٥	١٢٥	٦٢٥
مجـ س = ٩	مجـ س٢ = ٣٥	مجـ س٣ = ١٥٣	مجـ س٤ = ٧٠٧

∴ العزم الصفرى الاول (م١) = $\frac{\text{مجـ س}}{\text{ن}} = \frac{٩}{٣} = ٣$

العزم الصفرى الثانى (م٢) = $\frac{\text{مجـ س}^{٢}}{\text{ن}} = \frac{٣٥}{٣} = ١١,٦٧$

العزم الصفرى الثالث (م٣) = $\frac{\text{مجـ س}^{٣}}{\text{ن}} = \frac{١٥٣}{٣} = ٥١$

العزم الصفرى الرابع (م٤) = $\frac{\text{مجـ س}^{٤}}{\text{ن}} = \frac{٧٠٧}{٣} = ٢٣٥,٦٧$

أما الصيغة العامة للعزوم الصفرية فى حالة البيانات المبوبة (التوزيعات التكرارية) فهى :

$$م_{د} = \frac{مجـ س^{د} ك}{مجـ ك}$$

حيث د = ١ ، ٢، ٣ ، ٤ ، .... ، وبوضع د = ١ ، ٢ ، ٣ ، ٤ نحصل على العزوم الاربعة حول الصفر على التوالى كما يلى :

$$م_{١} = \frac{مجـ س ك}{مجـ ك} \qquad م_{٢} = \frac{مجـ س^{٢} ك}{مجـ ك}$$

$$م_{٣} = \frac{مجـ س^{٣} ك}{مجـ ك} \qquad م_{٤} = \frac{مجـ س^{٤} ك}{مجـ ك}$$

مثال (٢) :-

احسب العزوم الصفرية الاربعة الاولى للتوزيع التكرارى التالى:

فئات	١-	٣-	٥-٧
تكرارات	٣	٥	٢

يمكن حساب العزوم الصفرية الاربعة الاولى من الجدول التالى :

فئات	تكرارات (ك)	مراكز الفئات (س)	س ك	$س^{٢}$ك	$س^{٣}$ك	$س^{٤}$ك
١ -	٣	٢	٦	١٢	٢٤	٤٨
٣ -	٥	٤	٢٠	٨٠	٣٢٠	١٢٨٠
٥ - ٧	٢	٦	١٢	٧٢	٤٣٢	٢٥٩٢
المجموع	١٠	ــــ	٣٨	١٦٤	٧٧٦	٣٩٢٠

∴ العزم الصفرى الاول = ($م_{١}$) = مجـ س ك / مجـ ك = ٣٨ / ١٠ = ٣,٨

العزم الصفرى الثانى = ($م_{٢}$) = مجـ $س^{٢}$ ك / مجـ ك = ١٦٤ / ١٠ = ١٦,٤

العزم الصفرى الثالث = ($م_{٣}$) = مجـ $س^{٣}$ ك / مجـ ك = ٧٧٦ / ١٠ = ٧٧,٦

العزم الصفرى الرابع = ($م_{٤}$) = مجـ $س^{٤}$ك / مجـ ك = ٣٩٢٠ / ١٠ = ٣٩٢

يتضح من المثالين السابقين أن العزم الصفرى الأول سواء للبيانات الغير مبوبة أو المبوبة ما هو الا الوسط الحسابى وهو احد مقاييس النزعة المركزية الهامة ، كما يتبين لنا أننا لا نستطيع حساب العزوم الصفرية وأيضا العزوم المركزية كما سنرى من جداول تكرارية مفتوحة وذلك نتيجة اعتماد العزوم عند حسابها على مراكز الفئات .

ثانيا : العزوم المركزية (حول الوسط الحسابى (م)

The Moments About Mean

نفرض أن س متغير يأخذ القيم $س_١$ ، $س_٢$ ، $س_٣$ ، .... ، $س_ن$ فى عينة حجمها (ن) فلو رمزنا للعزم رقم (د) (العزم الدالى) حول الوسط الحسابى بالرمز (م) د فان :

الصيغة العامة للعزوم المركزية فى حالة البيانات الغير مبوبة هى:

$$\text{م}_{\text{د}} = \frac{\text{مجـ}\,(\text{س} - \bar{\text{س}})^{\text{د}}}{\text{ن}}$$

حيث د = ١ ، ٢ ، ٣ ، ... وبوضع د = ١ ، ٢ ، ٣ ، ٤ نحصل على العزوم المركزية الاربعة على التوالى كما يلى :

$$\text{م}_{١} = \frac{\text{مجـ}\,(\text{س} - \bar{\text{س}})}{\text{ن}}$$

$$\text{م}_{٢} = \frac{\text{مجـ}\,(\text{س} - \bar{\text{س}})^{٢}}{\text{ن}}$$

$$\text{م}_{٣} = \frac{\text{مجـ}\,(\text{س} - \bar{\text{س}})^{٣}}{\text{ن}}$$

$$\text{م}_{٤} = \frac{\text{مجـ}\,(\text{س} - \bar{\text{س}})^{٤}}{\text{ن}}$$

مثال (٢) :-

في مثال (١) أوجد العزوم المركزية الاربعة الاولى :

الحل :

يمكن حساب العزوم المركزية الاربعة من الجدول التالى :

حيث $\overline{س}$ = $م_١$ = ٣ علما بأن القيم الاصلية هى ١ ، ٣ ، ٥

(س − $\overline{س}$)	$(س - \overline{س})^٢$	$(س - \overline{س})^٣$	$(س - \overline{س})^٤$
١−٣=−٢	٤	−٨	١٦
٣−٣=صفر	صفر	صفر	صفر
٥−٣=٢	٤	٨	١٦
المجموع = صفر	٨	صفر	٣٢

∴ العزم المركزى الاول $م_١$ = $\frac{مجـ (س - \overline{س})}{ن}$ = $\frac{صفر}{٣}$ = صفر

∴ العزم المركزى الثانى $م_٢$ = $\frac{مجـ (س - \overline{س})^٢}{ن}$ = $\frac{٨}{٣}$ =٢,٦٧

∴ العزم المركزى الثالث $م_٣$ = $\frac{مجـ (س - \overline{س})^٣}{ن}$ = $\frac{صفر}{٣}$ = صفر

∴ العزم المركزى الرابع $م_٤$ = $\frac{مجـ (س - \overline{س})^٤}{ن}$ = $\frac{٣٢}{٣}$ = ١٠,٦٧

أما الصيغة العامة للعزوم المركزية فى حالة البيانات المبوبة هى:

$$م_د = \frac{مجـ (س - \overline{س})^د ك}{مجـ ك}$$

وبوضع د = ١ ، ٢ ، ٣ ، ٤ نحصل على العزوم المركزية الاربعة الاولى على النحو التالى :

$$م_١ = \frac{مجـ (س - \overline{س}) ك}{مجـ ك}$$

$$م_٢ = \frac{مجـ (س - \overline{س})^٢ ك}{مجـ ك}$$

$$م_٣ = \frac{مجـ (س - \overline{س})^٣ ك}{مجـ ك}$$

$$م_٤ = \frac{مجـ (س - \overline{س})^٤ ك}{مجـ ك}$$

مثال (٤) :-

فى مثال (٢) احسب العزوم المركزية الاربعة الاولى .

الحل

يمكن حساب العزوم المركزية الاربعة الاولى من الجدول التالى :

حيث س̄ = م١ = ٣.٨

فئات	تكرارات (ك)	مراكز الفئات (س)	(س - س̄)	(س - س̄) ك	(س - س̄)^{2}ك	(س - س̄)^{3}ك	(س - س̄)^{4}ك
١-	٣	٢	٢-٣,٨=-١,٨	-٥,٤	٩,٧٢	-١٧,٥	٣١,٥
٣-	٥	٤	٤-٣,٨=٠,٢	١	٠,٢	٠,٠٤	٠,٠٠٨
٥-٧	٢	٦	٦-٣,٨=٢,٢	٤,٤	٩,٦٨	٢١,٣	٤٦,٨٦
	١٠		صفر	صفر	١٩,٦	٣,٨٤	٧٨,٣٧

∴ العزم المركزى الاول (م١) = مجـ (س - س̄) ك / مجـ ك = صفر / ١٠ = صفر

∴ العزم المركزى الثانى (م٢) = مجـ (س - س̄)^{2} ك / مجـ ك = ١٩,٦ / ١٠ = ١,٩٦

∴ العزم المركزى الثالث (م٣) = مجـ (س - س̄)^{3} ك / مجـ ك = ٣,٨٤ / ١٠ = ٠,٢٨٤

∴ العزم المركزى الرابع (م٤) = مجـ (س - س̄)^{4}ك / مجـ ك = ٧٨,٣٧ / ١٠ = ٧,٨٣٧

ومما ينبغى ملاحظته من خلال المثالين السابقين أن العزم المركزى الاول ($م_{1}$) [سواء لبيانات غير مبوبة أو بيانات مبوبة] دائما يساوى الصفر وهذا راجع الى الخاصية الثالثة للوسط الحسابى والتى تقضى بأن مجموع انحرافات القيم عن وسطها الحسابى يساوى صفر أى أن : $م_{1}$ = صفر

كما يلاحظ أن العزم المركزى الثانى ($م_{2}$) ، [سواء لبيانات غير مبوبة أو بيانات مبوبة] ما هو الا التباين ($ع^{2}$) أى أن : $م_{2} = ع^{2}$

وبذلك يكون الانحراف المعيارى (ع) = $\sqrt{م_{2}}$

كذلك يمكننا الاستعانة بالعزوم سواء الصفرية أو المركزية فى حساب معامل الاختلاف على النحو التالى حيث أن :

$$\text{معامل الاختلاف} = \frac{ع}{\bar{س}} \times ١٠٠$$

$$\text{معامل الاختلاف (باستخدام العزوم الرياضية)} = \frac{\sqrt{م_{2}}}{مَ_{1}} \times ١٠٠$$

مثال (٥) :-

مستخدما أسلوب العزوم الرياضية اوجد معامل الاختلاف للمثالين ٣ ، ٤

الحل

فى مثال (٣)

$$\text{معامل الاختلاف} = \frac{\sqrt{\text{م}_2}}{\text{م}'_1} \times ١٠٠$$

$$= \frac{\sqrt{٢,٦٧}}{٣} \times ١٠٠ = ٥٤,٤٧\%$$

فى مثال (٤)

$$\text{معامل الاختلاف} = \frac{\sqrt{\text{م}_2}}{\text{م}'_1} \times ١٠٠$$

$$= \frac{\sqrt{١,٩٦}}{٣,٨} \times ١٠٠ = ٣٦,٨٤\%$$

يلاحظ أن تشتت مفردات الظاهرة الاولى (مثال ٣) أكبر من تشتت مفردات

الظاهرة الثانية (مثال ٤) .

(٥-١-١) : العلاقة بين العزوم الصفرية (مَ) والعزوم المركزية (م) :-

ذكرنا فيما سبق أن العزوم الرياضية تفيد في معرفة الخصائص المميزة للتوزيع التكراري ، حيث يمكن باستخدام العزوم معرفة أهم مقاييس النزعة المركزية مثل الوسط الحسابي ، ولاحظنا أن الوسط الحسابي ما هو إلا العزم الصفري الأول ، كذلك فإن حساب بعض مقاييس التشتت المطلقة مثل التباين (الانحراف المعياري) ، حيث لاحظنا أيضا أن التباين ما هو الا العزم المركزي الثاني علاوة على امكانية الاستعانة بهذه العزوم في دراسة خاصيتي التماثل والاعتدال للتوزيعات التكرارية وهذا ما سيتضح في أجزاء لاحقة من هذا الباب .

ولما كانت طريقة حساب العزوم المركزية صعبة ومعقدة خاصة اذا كان الوسط الحسابي ($\overline{س}$) يحتوي على كسور مما يجعل انحرافات القيم أو مراكز الفئات يحتوي على كسور أيضا ، لذلك فقد وجد أنه من الأفضل ولدواعي السهولة والدقة أن يتم حساب العزوم المركزية بدلالة العزوم الصفرية ، وفيما يلي العلاقات المختلفة بين كل من العزوم الصفرية والعزوم المركزية (سوف نكتفي باثبات هذه العلاقات في حالة البيانات الغير مبوبة فقط) .

نعلم أن العزم المركزي الاول :

$$م_1 = \frac{مجـ (س - \overline{س})}{ن} = \frac{1}{ن} (مجـ س - \overline{س})$$

$$= \frac{\text{مجـ س}}{\text{ن}} - \frac{\overline{\text{س}}}{\text{ن}}$$

$$= \text{م}'_{١} - \text{م}'_{١} = \text{صفر}$$

وهذا ما يتفق ونتائج مثالى ٣ ، ٤ حيث $\text{م}'_{١}$ = صفرا دائما اما بالنسبة للعزم المركزى الثانى فإن :

$$\text{م}_{٢} = \frac{\text{مجـ}(\text{س} - \overline{\text{س}})^{٢}}{\text{ن}} = \frac{١}{\text{ن}} \text{ مجـ}(\text{س}^{٢} - ٢\text{س}\,\overline{\text{س}} + \overline{\text{س}}^{٢})$$

$$= \frac{١}{\text{ن}} \text{ مجـ س}^{٢} - ٢\overline{\text{س}} \frac{\text{مجـ س}}{\text{ن}} + \frac{١}{\text{ن}} \text{ن}\,\overline{\text{س}}^{٢}$$

$$= \frac{١}{\text{ن}} \text{ مجـ س}^{٢} - ٢\overline{\text{س}}^{٢} + \overline{\text{س}}^{٢}$$

$$= \text{م}'_{٢} - (\text{م}'_{١})^{٢}$$

وبنفس الاسلوب السابق يمكن اثبات أن :

$$\text{م}_{٣} = \text{م}'_{٣} - ٣\text{م}'_{٢}\,\text{م}'_{١} + ٢(\text{م}'_{١})^{٣}$$

$$\text{م}_{٤} = \text{م}'_{٤} - ٤\text{م}'_{٣}\,\text{م}'_{١} + ٦\text{م}'_{٢}(\text{م}'_{١})^{٢} - ٣(\text{م}'_{١})^{٤}$$

وقد سبق أن ذكرنا أن هناك تصحيحا خاصا بالعزم المركزى الثانى (التباين) والمعروف بتصحيح شبرد ويأخذ الصورة التالية :

$$(\text{م}_{٢})\text{ المصحح} = \text{م}_{٢} - \frac{\text{ب}^{٢}}{١٢}$$

ومن المفيد إضافة أن العزم المركزى الثالث لا يحتاج الى تصحيح ، فى حين أن العزم المركزى الرابع يمكن تصحيحه على النحو :

$$(\text{م}_{٤})\ \text{المصحح} = \text{م}_{٤} - \frac{\text{ب}^{٢}}{١٢}\ \text{م}_{٢} + \frac{٧}{٢٤٠}\ \text{ب}^{٤}$$

وهذه التصحيحات تقتصر فقط على الحالات التى يكون فيها منحنى التوزيع مستمرا ويقابل طرفاه محور السينات وأن تكون أطوال فئاته متساوية (توزيع منتظم) .

مثال (٦) :-

أ- فى مثال (١) استخدم العلاقة بين العزوم المركزية والصفرية للتأكد من صحة النتائج .

ب- فى مثال (٤) المطلوب تصحيح العزوم المركزية ($\text{م}_{٢}$ ، $\text{م}_{٤}$) للتوزيع التكرارى .

الحل :

أ- باستخدام بيانات مثال (١) والعلاقة بين العزوم المركزية والصفرية نجد أن :

$$\text{م}_{٢} = \text{م}'_{٢} - (\text{م}'_{١})^{٢}$$

$$= ١١,٦٧ - (٣)^{٢} = ٢,٦٧$$

مـ٣ = م٣َ − ٣م٢َ م١َ + ٢ (م١َ)٣

= ٥١ − ٣ × ١١,٦٧ × ٣ + ٢ (٣)٣ ≈ صفر

مـ٤ = م٤َ − ٤ م٣َ م١َ + ٦ م٢َ (م١َ)٢ − ٣ (م١َ)٤

= ٢٣٥,٦٧ − ٤ × ٥١ × ٣ + ٦ × ١١,٦٧ × (٣)٢ − ٣ (٣)٤

= ١٠,٨٥

وهى تقريبا نفس نتائج العزوم المركزية المتحصل عليها من مثال (٣) [الطريقة المباشرة] .

ب− من نتائج مثال (٤) يمكن تصحيح كل من العزوم المركزية مـ٢ ، مـ٤ على النحو التالى :

مـ٢ المصحح = مـ٢ − ب٢ / ١٢

= ١,٩٦ − (٢)٢ / ١٢ =

= ١,٩٦ − ٠,٣٤ =١,٦٢

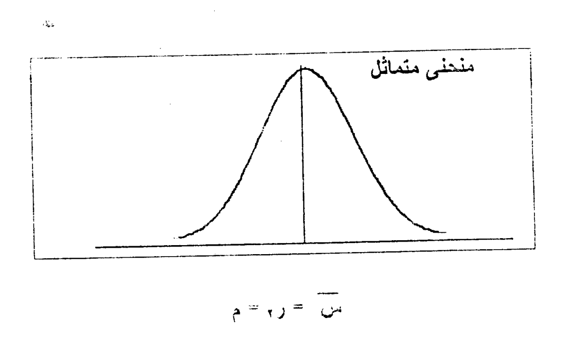
مـ٤ المصحح = مـ٤ − ب٢ / ١٢ مـ٢ + ٧ / ٢٤٠ ب٤

= ٧,٨٣٧ − (٢)٢ / ١٢ × ١,٩٦ + ٧ / ٢٤٠ (٢)٤

= ٧,٨٣٧ − ٠,[illegible]

(٥ – ٤) الالتواء Skewness :-

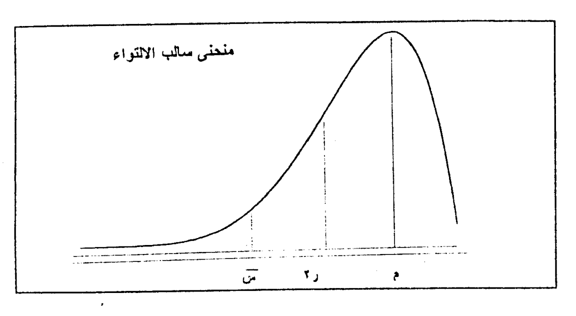
ذكرنا فيما سبق أن التوزيعات التكرارية وايضا المنحنيات التكرارية التي تمثلها قد تكون متماثلة كما قد تكون ملتوية أيضا . وعلمنا أن التوزيع المتماثل هو ذلك المنحنى الذي يقسمه محور التماثل الى قسمين متساويين ومتطابقين وأن التكرارات تتزايد أو تتناقص على جانبي محور التماثل بنفس الدرجة .



$\overline{س} = ر_٢ = م$

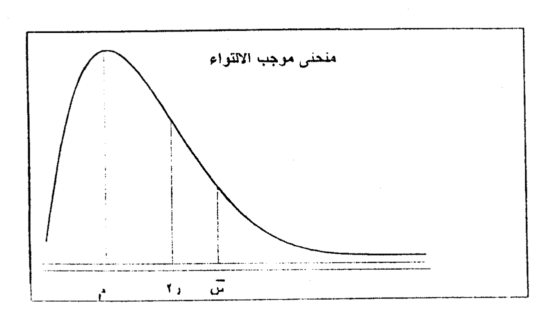
في حالة التوزيع المتماثل يلاحظ أن $\overline{س} = ر_٢ = م$

أما التوزيع سالب الالتواء فهو ذلك التوزيع الذي يزداد تركز المفردات فيه وتزداد درجة التركز في الاجزاء العليا للتوزيع ويمتد ذيل المنحنى التكراري الى يسار التوزيع .



فى حالة التوزيع سالب الالتواء يلاحظ أن س̄ < ر٢ < م

وعلى العكس من ذلك يكون التوزيع موجب الالتواء حين يزداد تركزالمفردات ويزداد التجانس عند الفئات الدنيا ، للتوزيع ويمتد ذيل المنحنى التكرارى الى يمين التوزيع .



فى حالة التوزيع موجب الالتواء يلاحظ أن $\overline{س} > ر_٢ > م$

وسبق أن رأينا فى الباب الثالث عند استعراض العلاقة بين مقاييس النزعة المركزية الثلاثة (الوسط والوسيط والمنوال) أن هذه العلاقة [illegible] خصائص التوزيع التكرارى من حيث التماثل (الالتواء) وقد [illegible] العلاقات الثلاث لاستنتاج عدة صيغ رياضيه تسمى بمعاملات الالتواء [illegible] :

١ – معامل بيرسون للالتواء : وهو عبارة عن صيغتين هما :

$$ت_١ = \frac{\overline{س} - م}{ع} \quad \text{(باستخدام الوسط والمنوال)}$$

حيث يكون $-٣ \leq ت_١ \leq +٣$ دائما .

$$ت_٢ = \frac{٣ (\overline{س} - ر_٢)}{ع} \quad \text{(باستخدام الوسط والوسيط)}$$

حيث يكون $-٣ \leq ت_٢ \leq +٣$ دائما

٢ – معامل بولى للالتواء

واضح أن كلا الصيغتين السابقتين لا يمكن حسابهما للتوزيعات [illegible]

[illegible]

حسابية طويلة ، لذلك يفضل استخدام صيغة بولـى فـى حالـة التوزيعات التكراريـة المفتوحة والتى تعتمد على الوسيط ($ر_٢$) وشبيهات الوسيط ($ر_١$ ، $ر_٣$) حيث :

$$ت_٣ = \frac{(ر_٣ - ر_٢) - (ر_٢ - ر_١)}{(ر_٣ - ر_١)}$$

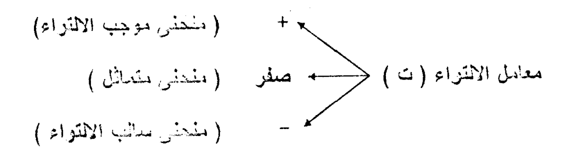
باستخدام ($ر_١$ ، $ر_٢$ ، $ر_٣$) حيث $-١ \leq ت_٣ \leq +١$ دائما

٣- معامل الالتواء باستخدام العزوم المركزية

كذلك يمكن قياس الالتواء باستخدام العزوم الرياضية وذلك بالاعتماد على حساب كل من قيمة العـزم المركـزى الثـالث ($م_٣$) وايضـا قيمـة العزم المركـزى الثانى ($م_٢$) ويأخذ معامل الالتواء الصيغة التالية :

$$ت_٤ = \frac{م_٣}{\sqrt{م_٢^{٣}}}$$

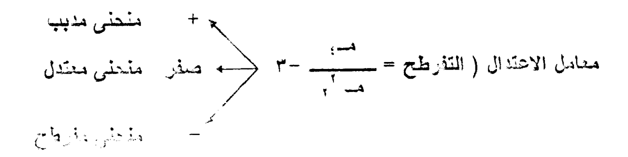
عموما يمكن وضع شكل عام لمفهوم خاصية الالتواء بعد استخدام أى من المعاملات السابقة ثم مقارنة الناتج بالصفر وذلك لاستنتاج شكل منحنى التوزيع على النحو التالى :



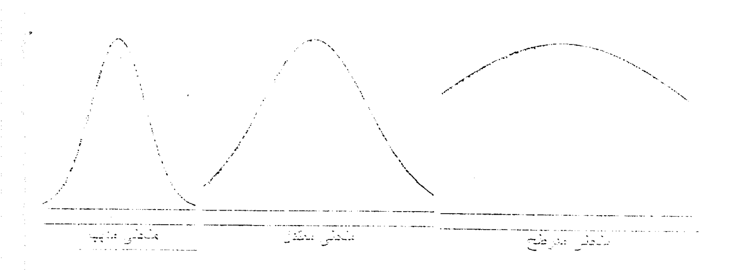
وجدير بالملاحظة أنه عندما تصل قيمة معامل الالتواء الى (± ١):ل ذلك على أن التوزيع موضوع الدراسة شديد الالتواء .

<u>(٥ – ٣) التفرطح *Kurtosis* :–</u>

عادة ما نلاحظ عند مقارنة المنحنيات وحيدة القمة أنها متساوية من حيث (القيمة المتوسطة والتشتت والالتواء) الا أننا قد نجدها مختلفة من حيث شكل القمة ، حيث قد نجد أن بعض هذه القمم أكثر تدببا أو تفرطحا من بعض القمم الاخرى . ولذلك فقد لزم الامر ايجاد مقياس جديد لقياس خاصية التفرطح ويقوم هذا المقياس على الاساس التالى :



حيث يمكن رسم أشكال منحنى التوزيع حسب درجة تفرطحه على النحو التالى :



<u>* ملاحظات حول الالتواء والتفرطح :-</u>

- معاملات الالتواء والتفرطح كلها مقاييس نسبية ليس لها تمييز ولذلك تصلح هذه المعاملات لمقارنة توزيعات تختلف فى وحدات قياسها كالاطوال والاعمار مثلا .
- يلاحظ أن قيم معاملات الالتواء السابقة قد تختلف بالنسبة للتوزيع الواحد وذلك بسبب اختلاف تعريف كل منها . وهذا يعنى ضرورة استخدام نفس المعامل عند مقارنة درجة الالتواء لاكثر من توزيع .

- تتوقف اشارة اى من معاملات الالتواء السابقة على اشارة البسط وذلك لأن اشارة المقام دائما موجبة باستثناء حالة وحيدة كون ع أو ($ر_٢ - ر_١$) = صفر كما فى حالة التجانس الكامل للبيانات موضوع الدراسة.

<u>مثال (٧) :-</u>

ادرس خاصيتى الالتواء (التماثل) والتفرطح (الاعتدال) للتوزيع التكرارى الموضح بمثال (٤)

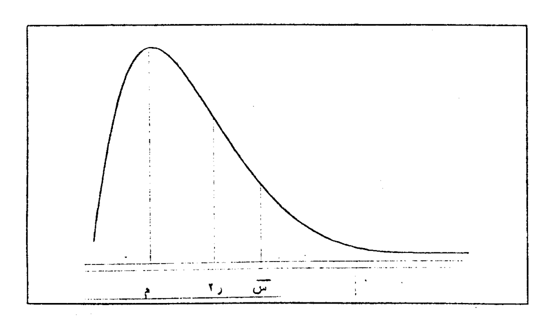
<u>الحل :</u>

(أ) بالنسبة لدراسة خاصية الالتواء يلاحظ أن :

$$\text{معامل الالتواء (باستخدام العزوم الرياضية)} = \frac{م_٣}{\sqrt{م_٢^{٣}}}$$

$$= \frac{٠,٣٨٤}{\sqrt{(١,٦٦)^{٣}}} = + ٠,١٤$$

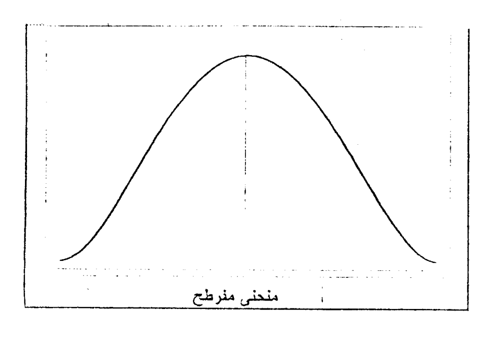
وهذا يدل على أن توزيع الظاهرة موجب الالتواء ويأخذ منحنى هذا التوزيع الشكل التالى :



$$\text{معامل التفرطح} = \frac{\mu_4}{\mu_2^2} - 3$$

$$= \frac{7{,}837}{(1{,}96)^2} - 3 = -0{,}959$$

مما يدل على أن توزيع الظاهرة مفرطح ويأخذ منحنى هذا التوزيـع الشـكل التالى :


منحنى مفرطح

الباب السادس

الارتباط الخطى

Linear Correlation

تكلمنا فى الابواب السابقة عن المفاهيم الاحصائية الخاصة بوصف مجموعة من قيم متغير واحد أو ظاهرة واحدة وعن الخصائص الاساسية للتوزيع التكرارى لهذه الظاهرة وذلك عن طريق التعرف على طرق حساب وكيفية الاستفادة من بعض المقاييس الاحصائية مثل مقاييس النزعة المركزية ومقاييس التشتت والالتواء والتفرطح وغيرها .

و فى هذا الباب سوف ندرس نوعا آخر من أنواع التحليل الاحصائى الخاصة بالعلاقة بين ظاهرتين ، فإذا أخذنا مثلا مجموعة من الطلبة فى إحدى الكليات وسألنا كلا منهم عن وزنه وطوله ، لتجمع لدينا مجموعة من البيانات أو القرارات عن متغيرين أو ظاهرتين هما (ظاهرة الطول وظاهرة الوزن) ومن الطبيعى وجود علاقة بين هاتين الظاهرتين بمعنى أن وزن الطالب يتبع فى تغيره – بصفة عامة – الطول ، فكلما زاد طول الطالب زاد وزنه . وتسمى مثل هذه العلاقة بين الظاهرتين بالارتباط ، وعادة ما يقال على مثل هذان المتغيران بأنهما مرتبطان – أى أن الارتباط بمعناه البسيط هو العلاقة بين متغيرين أو أكثر .

ويبين الارتباط درجة العلاقة بين المتغيرين كما يبين شكل هذه العلاقة واتجاهها أيضا . فإذا كان التغير فى احد المتغيرين يتبعه تغير بالزيادة أو النقصان فى المتغير الاخر فى نفس الاتجاه ، فى هذه الحالة يمكننا القول بأن الارتباط بين المتغيرين طردى (موجب) . أما اذا كان التغير فى المتغيرين فى اتجاهين متضادين - بمعنى أن زيادة أحدهما تؤدى الى نقص الاخر والعكس صحيح ، فإننا نقول أن الارتباط فى هذه الحالة بين المتغيرين سالب (عكس) .

وما ينبغى التنوية عنه أن وجود ارتباط بين ظاهرتين لا يعتبر دليل على أن أحد الظاهرتين تحدث نتيجة لحدوث الاخرى ولا ينشأ إلا بسببه ، إذ قد يكون هناك مؤثر أو عدة مؤثرات تؤثر فى الظاهرتين معا ، بحيث تجعل تغير إحداهما يظهر كما لو كان نتيجة لتغير الظاهرة الاخرى .

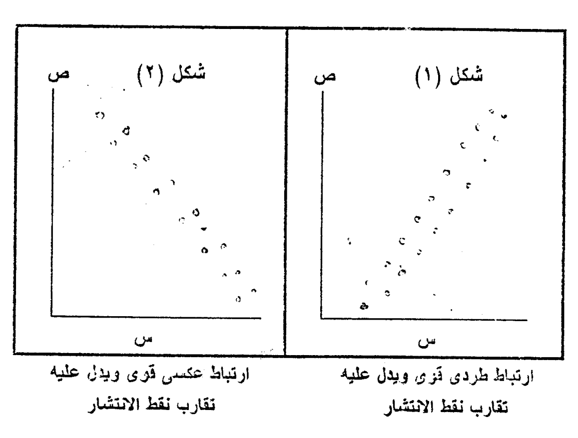
__*Scatter Diagram* شكل الانتشار__

وعموما اذا كان لدينا عدد (ن) من أزواج القيم الخاصة بالمتغيرين س ، ص أى :

(س$_{١}$ ، ص$_{١}$) ، (س$_{٢}$ ، ص$_{٢}$) ، .... (س$_{ن}$ ، ص$_{ن}$)

فانه يمكن استخدام التمثيل البيانى (عن طريق رسم النقط التى تمثل المتغيرين) فى دراسة العلاقة بين هذين المتغيرين وهو ما يعرف بشكل الانتشار . حيث يوضح لنا هذا الشكل درجة العلاقة بين المتغيرين واتجاه هذه العلاقة ، فإذا

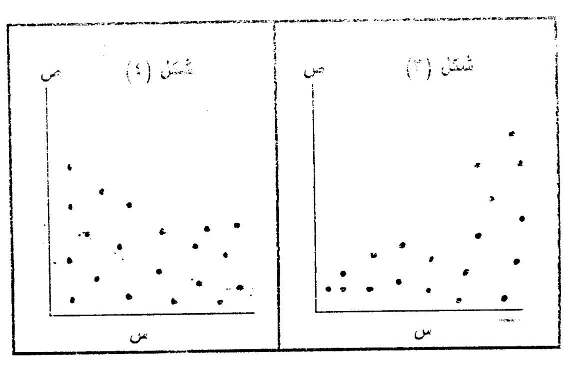
كانت نقط شكل الانتشار متقاربة من بعضها البعض دل ذلك على وجود ارتباط طردى (أو عكس) قوى ، أما اذا كانت النقط فى شكل الانتشار بعيدة عن بعضها البعض فإن هذا يعنى أن الارتباط بين المتغيرين س ، ص ضعيف ، وشكلا (١) ، (٢) يوضحان ارتباطا طرديا وارتباطا عكسيا قويا بين المتغيرين (س ، ص) بينما يظهر الشكلان (٣) ، (٤) ارتباطا طرديا ضعيفا وارتباطا عكسيا ضعيفا على الترتيب .



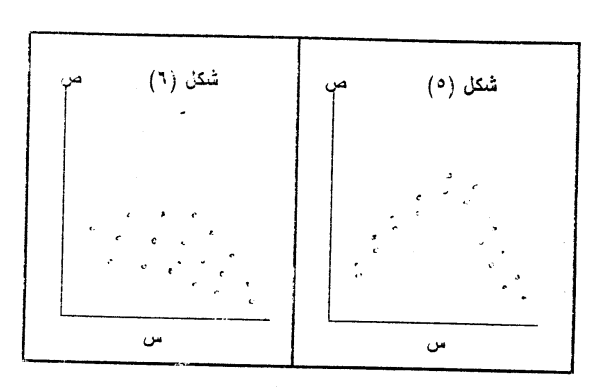
ارتباط طردى قوى ويدل عليه تقارب نقط الانتشار

ارتباط عكسى قوى ويدل عليه تقارب نقط الانتشار

اما الشكلان التاليان (٣) ، (٤) فيدل الاول على ارتباط طردى ضعيف و هذا واضح من تباعد نقط شكل الانتشار ، بينما يدل الشكل الثانى على ارتباط طردى ضعيف ايضا و لكن فى الاتجاه العكسى .



وبالاضافة الى ما سبق فإن شكل الانتشار يوضح لنا ما اذا كانت العلاقة بين المتغيرين يمثلها خط مستقيم أو منحنى . فإذا كانت نقط شكل الانتشار تتجمع حول مستقيم كما هو واضح فى الاشكال (١) ، (٢) ، (٣) ، (٤) فيمكننا القول بأن العلاقة بين المتغيرين (س) ، (ص) علاقة خطية أما اذا كانت نقط الانتشار تتجمع حول منحنى فإن العلاقة بين المتغيرين (س) ، (ص) تسمى علاقة غير خطية كما هو واضح من شكل (٥) التالى علاوة على وجود بعض الحالات التى لا يوجد فيها ارتباط على الاطلاق كما هو واضح من شكل (٦) التالى ايضا .



كما يلاحظ أنه اذا وقعت جميع نقط شكل الانتشار علـى خط مستقيم ، فى هذه الحالة يمكننا القول بأن العلاقة بين المتغـيرين (س) ، (ص) هى علاقـة تامـة (طردية أو عكسية) كما هو واضح من الشكلان (٧) ، (٨) :

شكل (٨)
ص
س
شكل (٧)
ص
س

واتضح من استعراض أشكال الانتشار السابقة أنها تصف لنا فقط العلاقة بين المتغيرين (س) ، (ص) من حيث القوة والضعف , كذلك كون هذه العلاقة طردية أو عكسية . الا أنها لا تعطينا المقياس الرقمى المحدد لدرجة هذه العلاقة لذلك فقد ظهرت الحاجة لمقياس رقمى يصور لنا درجة العلاقة بين المتغيرين وهذا ما سنطلق عليه " معامل الارتباط (ر) " والذى تقتصر فائدته على اثبات وجود علاقة من عدمها بالاضافة الى قياس درجتها ، إلا أن هذا المعامل لا يمكننا من ايجاد صورة رياضية بين المتغيرين بحيث يمكن استخدامها فى وصف العلاقة بينهما وأيضا استخدامها فى التنبؤ بما سيحدث للعلاقة بين المتغيرين فى المستقبل.

وعموما فإن قيمة معامل الارتباط تنحصر دائما بين -١ ، +١ أى أن :

$$-١ \leq ر \leq +١$$

وهذه المتباينة تحتوى على حدود نوعين من الارتباط :

الاول : ارتباط طردى وفيه صفر ≤ ر ≤ +١

الثانى : ارتباط عكسى وفيه -١ ≤ ر ≤ صفر

فإذا كانت ر = +١ فنقول أن الارتباط طردى تام

أما اذا كانت ر = -١ فنقول أن الارتباط عكسى تام

وعادة ما تقاس قوة الارتباط بمقدار قرب (ر) من -١ أو +١

وفيما يلى استعراض لأهم مقاييس الارتباط فى الحالات المختلفة :

(٦-١) معامل ارتباط بيرسون *Pearson Correlation Coefficient*

أولا معامل ارتباط بيرسون للبيانات الغير مبوبة :

تقاس درجة الارتباط بين المتغيرات (الظواهر).بواسطة معامل الارتباط (ر) . فنفرض أن لدينا مجموعة مكونة من (ن) طالبا بإحدى الجامعات وحصلنا من هؤلاء الافراد على بيانات عن قيم ظاهرتين (مثل الطول والوزن) ورمزنا للظاهرة الاولى بالرمز (س) وللظاهرة الثانية (ص) فتكون البيانات التى لدينا على الصورة : الظاهرة الاولى (س) : $\text{س}_{١} ، \text{س}_{٢} ، \text{س}_{٣} ، \ldots\ldots ، \text{س}_{\text{ن}}$

الظاهرة الثانية (ص) : $\text{ص}_{١} ، \text{ص}_{٢} ، \text{ص}_{٣} ، \ldots\ldots ، \text{ص}_{\text{ن}}$

ويأخذ معامل ارتباط بيرسون للبيانات الغير مبوبة إحدى الصيغ الاتية :

$$\text{ر} = \frac{\text{تغا} (\text{س} ، \text{ص})}{\sqrt{\text{تبا} (\text{س}) \times \text{تبا} (\text{ص})}}$$

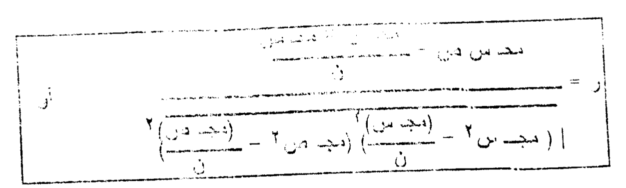
حيث $\text{تغا} (\text{س} ، \text{ص}) = \text{تغاير} (\text{س} ، \text{ص}) = \frac{١}{\text{ن}} \text{مجـ} (\text{س} - \overline{\text{س}}) (\text{ص} - \overline{\text{ص}})$

أما

$\text{تبا} (\text{س}) = \text{تباين} (\text{س}) = \frac{١}{\text{ن}} \text{مجـ} (\text{س} - \overline{\text{س}})^{٢}$ أما

$\text{تبا} (\text{ص}) = \text{تباين} (\text{ص}) = \frac{١}{\text{ن}} \text{مجـ} (\text{ص} - \overline{\text{ص}})^{٢}$

$$\therefore \text{ر} = \frac{\text{مجـ} (\text{س} - \overline{\text{س}}) (\text{ص} - \overline{\text{ص}})}{\sqrt{\text{مجـ} (\text{س} - \overline{\text{س}})^{٢} \times \text{مجـ} (\text{ص} - \overline{\text{ص}})^{٢}}}$$



$$ر = \frac{\frac{1}{ن}\,\text{مجـ}\,س\,ص - \bar{س}\,\bar{ص}}{ع_س\,ع_ص}$$

$\bar{س}$: يمثل الوسط الحسابي للظاهرة س = $\frac{\text{مجـ}\,س}{ن}$

$\bar{ص}$: يمثل الوسط الحسابي للظاهرة ص = $\frac{\text{مجـ}\,ص}{ن}$

ع س : تمثل الانحراف المعياري للظاهرة س = $\sqrt{\frac{1}{ن}\,\text{مجـ}\,س^{2} - (\bar{س})^{2}}$

ع ص : يمثل الانحراف المعياري للظاهرة ص = $\sqrt{\frac{1}{ن}\,\text{مجـ}\,ص^{2} - (\bar{ص})^{2}}$

مثال (١) :

أوجد معامل الارتباط بين أطوال وأوزان ٨ طلاب من طلبة إحدى الجامعات من البيانات الآتية :

الطول (سم) :	١٦٤	١٥٢	١٨٤	١٦٤	١٧٦	١٥٦	١٦٨	١٦٤
الوزن (كجم)	٥٢	٤٠	٦٠	٥٢	٦٠	٤٢	٥٠	٥٢

الحل :

نفرض أن (س) ترمز لقيم الطول وأن (ص) ترمز لقيم ظاهرة الوزن وبالتالى يمكننا تكوين الجدول التالى :

س	ص	س٢	ص٢	س ص
١٦٤	٥٢	٢٦٨٩٦	٢٧٠٤	٨٥٢٨
١٥٢	٤٠	٢٣١٠٤	١٦٠٠	٦٠٨٠
١٨٤	٦٠	٣٣٨٥٦	٣٦٠٠	١١٠٤٠
١٦٤	٥٢	٢٦٨٩٦	٢٧٠٤	٨٥٢٨
١٧٦	٦٠	٣٠٩٧٦	٣٦٠٠	١٠٥٦٠
١٥٦	٤٢	٢٤٣٣٦	١٧٦٤	٦٥٥٢
١٦٨	٥٠	٢٨٢٢٤	٢٥٠٠٠-	٨٤٠٠
١٦٤	٥٢	٢٦٨٩٦	٢٧٠٤	٨٥٢٨
مجـ =١٣٢٨	٤٠٨	٢٢١١٨٤	٢١١٧٦	٦٨٢١٦

(أ) الوسط الحسابى للظاهرتين

$$\bar{س} = \frac{مجـ\ س}{ن} = \frac{١٣٢٨}{٨} = ١٦٦$$

$$\bar{ص} = \frac{مجـ\ ص}{ن} = \frac{٤٠٨}{٨} = ٥١$$

(ب) الانحراف المعياري لكل من المتغيرين :

$$\text{ع}_{\text{س}} = \sqrt{\frac{\text{مجـ س}^{2}}{\text{ن}} - \bar{\text{س}}^{2}} = \sqrt{\frac{\text{٢٢١١٨٤}}{\text{٨}} - (\text{١٦٦})^{2}}$$

$$= \sqrt{\text{٢٧٦٤٨} - \text{٢٧٥٥٦}} = \sqrt{\text{٩٢}}$$

$$\text{ع}_{\text{ص}} = \sqrt{\frac{\text{مجـ ص}^{2}}{\text{ن}} - \bar{\text{ص}}^{2}} = \sqrt{\frac{\text{٢١١٧٦}}{\text{٨}} - (\text{٥١})^{2}}$$

$$= \sqrt{\text{٢٦٤٧} - \text{٢٦٠١}} = \sqrt{\text{٤٦}}$$

وعلى ذلك فمعامل الارتباط هو :

$$\text{ر} = \frac{\frac{\text{مجـ س ص}}{\text{ن}} - \bar{\text{س}}\ \bar{\text{ص}}}{\text{ع}_{\text{س}}\ \text{ع}_{\text{ص}}} = \frac{\frac{\text{٦٨٢١٦}}{\text{٨}} - \text{١٦٦} \times \text{٥}}{\sqrt{\text{٩٢}}\ \sqrt{\text{٤٦}}} = \text{٠,٩٤}$$

كما يمكننا استخدام الصيغة التالية مباشرة لايجاد معامل الارتباط على النحو التالي :

$$\text{ر} = \frac{\text{مجـ س ص} - \frac{\text{مجـ س} \times \text{مجـ ص}}{\text{ن}}}{\sqrt{\left(\text{مجـ س}^{2} - \frac{(\text{مجـ س})^{2}}{\text{ن}}\right)\left(\text{مجـ ص}^{2} - \frac{(\text{مجـ ص})^{2}}{\text{ن}}\right)}}$$

$$= \frac{٦٨٢١٦ - \frac{١٣٢٨ \times ٤٠٨}{٨}}{\sqrt{\left(٢٢١١٨٤ - \frac{(١٣٢٨)^{٢}}{٨}\right)\left(٢١١٧٦ - \frac{(٤٠٨)^{٢}}{٨}\right)}} = ٠,٩٤$$

وأيا كانت الصيغة المستخدمة فى حساب معامل الارتباط بين ظاهرتى الطول والوزن فان الناتج السابق لمعامل الارتباط يعنى وجود ارتباط طردى قوى وليس تام بحيث أنه لو رسمنا شكل الانتشار للمثال السابق لتجمعت النقاط حول الخط ولكنها لا تقع كلها على الخط المستقيم .

<u>مثال (٢) :</u>

بفرض أنه تجمعت لديك البيانات الخاصة بدخول خمسة من العمال وأيضا سنوات خبرتهم فى العمل – المطلوب حساب معامل الارتباط (ر) .

الدخل فى اليوم(بالجنيه)	٤	٨	١٢	١٦	٢٠
سنوات الخبرة	١	٢	٣	٤	٥

الحل

يمكن استخدام اى من الصيغ الثلاثة السابقة فى حساب معامل الارتباط بين الظاهرتين وذلك بتكوين الجدول التالى :-

ص	س	ص٢	س٢	س ص
٤	١	١٦	١	٤
٨	٢	٦٤	٤	١٦
١٢	٣	١٤٤	٩	٣٦
١٦	٤	٢٥٦	١٦	٦٤
٢٠	٥	٤٠٠	٢٥	١٠٠
مجـ = ٦٠	١٥	٨٨٠	٥٥	٢٢٠

$$ر = \frac{\text{مجـ س ص} - \frac{\text{مجـ س} \times \text{مجـ ص}}{ن}}{\sqrt{\left(\text{مجـ س}^{٢} - \frac{(\text{مجـ س})^{٢}}{ن}\right)\left(\text{مجـ ص}^{٢} - \frac{(\text{مجـ ص})^{٢}}{ن}\right)}}$$

$$= \frac{٢٢٠ - \frac{٦٠ \times ١٥}{٥}}{\sqrt{\left(٨٨٠ - \frac{(٦٠)^{٢}}{٥}\right)\left(٥٥ - \frac{(١٥)^{٢}}{٥}\right)}} = \frac{٢٢٠ - ١٨٠}{\sqrt{١٦٠ \times ١٠}} = \frac{٤٠}{٤٠} = ١$$

وهذا يعنى أن العلاقة طردية تامة بين المتغيرين (الدخل وسنوات الخبرة) بحيث اذا رسمنا شكل انتشار لتمثيل هذه العلاقة بيانيا فسوف نلاحظ تجمع هذه القيم على خط يصنع زاوية حاده مع المحور الافقى .

مثال (٢) :

بفرض أنه توافرت لديك القيم الآتية عن الظاهرتين س ، ص

الظاهرة الأولى (س) :

الوسط الحسابى = ٥

مجموع مربعات قيم الظاهرة = ٢٢٣

مجموع مربعات انحرافات القيم عن وسطها الحسابى = ٤٨

الظاهرة الثانية (ص) :

مجموع القيم = ٥٦

مجموع مربعات انحرافات القيم عن وسطها الحسابى = ٧٢

المطلوب : حساب معامل ارتباط بيرسون بين الظاهرتين س ، ص وذلك اذا علمت أن مجـ س ص = ٣١٦

الحل :

من المعلومات المتوفرة فى التمرين عن الظاهرة الأولى (س) أن :

الوسط الحسابى = $\bar{س}$ = ٥

، مجموع مربعات القيم = مجـ $س^{٢}$ = ١٢٣

، مجموع مربعات القيم عن وسطها الحسابى = مجـ $(س - \bar{س})^{٢}$ = ٤٨

وحيث أن :

مجـ (س − $\bar{س}$)^2 = مجـ س^2 − ن $\bar{س}^2$

∴ ٤٨ = ٢٢٣ − ن (٥)^2

٢٥ ن = ٢٢٣ − ٤٨

∴ ن = $\frac{١٧٥}{٢٥}$ = ٧

كما يلاحظ من البيانات المتوفرة عن الظاهرة الثانية (ص) أن مجموع القيم = مجـ ص = ٥٦

∴ الوسط الحسابى ($\bar{ص}$) = $\frac{مجـ ص}{ن}$ = $\frac{٥٦}{٧}$ = ٨

وأن مجـ س ص = ٣١٦

∴ مجـ س$'$ ص$'$ = مجـ س ص − ن $\bar{س}$ $\bar{ص}$

= ٣١٦ − ٧ × ٥ × ٨ = ٣٦

∴ ر = $\frac{مجـ س' ص'}{\sqrt{مجـ س'^2 مجـ ص'^2}}$

= $\frac{٣٦}{\sqrt{٤٨ \times ٧٢}}$ = ٠,٦١

وهذا يعنى أن العلاقة بين الظاهرتين س ، ص طردية قوية .

أما بالنسبة للبيانات المبوبة فاننا لا نحتاج الى رسم شكل الانتشار لاستطلاع العلاقة بين المتغيرين (س ، ص) لتقرر أى صيغة سوف نستخدمها وذلك لأن البيانات فى هذه الحالة سوف تكون معروضة فى صورة جدول تكرارى مزدوج أى جدول تكرارى ثنائى تخصص فيه الصفوف لفئات أحد المتغيرين والاعمدة لفئات المتغير الاخر . ويتحدد موقع كل نقطة - بقيمة من (س) تناظرها قيمة من (ص) - على أساس فئة من فئات (س) وفئة تناظرها من (ص) وهذه هى النقط المشتركة . والجدول التكرارى المزدوج بصورته هذه يمكن أن ينظر اليه على أنه صورة أخرى للشكل الانتشارى محوراه الافقى والرأسى هما فئات (س) للصفوف وفئات (ص) للأعمدة والنقط تعبر عنها التكرارات المشتركة وبالتالى فانه يمكن ملاحظة شكل انتشار هذه التكرارات وتقدير شكل ودرجة العلاقة بين المتغيرين .

ويمكن حساب معامل الارتباط من الجداول التكرارية باستخدام الصيغة الاتية :

$$\text{ر} = \frac{\text{مجـ س ص ك} - \dfrac{\text{مجـ س ك} \times \text{مجـ ص ك}}{\text{مجـ ك}}}{\sqrt{\left(\text{مجـ س}^{2}\text{ ك} - \dfrac{(\text{مجـ س ك})^{2}}{\text{مجـ ك}}\right)\left(\text{مجـ ص}^{2}\text{ ك} - \dfrac{(\text{مجـ ص ك})^{2}}{\text{مجـ ك}}\right)}}$$

حيث :

س ، ص تمثل مراكز فئات المتغيرين س ، ص على الترتيب .

ك : تمثل التكرارات الخاصة بكل فئة .

ولحساب معامل الارتباط (ر) من جدول مزدوج يراعى اجراء ما يلى :

١- تحسب كل من (مجـ س ك) ، (مجـ $س^٢$ ك) من التوزيع الهامشى للمتغير (س) وذلك كما ذكرنا ، إما باستخدام مراكز الفئات (س) أو الفروق البسيطة (ح س) أو الفروق المعدلة (ح/ س) .

٢- كذلك نحسب كل من (مجـ ص ك) ، (مجـ $ص^٢$ ك) من التوزيع الهامشى للمتغير (ص) وذلك ، إما باستخدام مراكز الفئات (ص) أو الفروق البسيطة (ح ص) أو الفروق المعدلة (ح/ ص)

٣- أما بالنسبة للمقدار (مجـ س ص ك) فيمكن الحصول عليه من جدول التوزيع المزدوج للمتغيرين س ، ص إما باستخدام مراكز الفئات والتكرارات المناظرة فى الجدول المزدوج أو باستخدام الفروق البسيطة أو الفروق المعدلة كما ورد فى كل من (١) ، (٢) .

مثال (٤) :

الجدول التكراري المزدوج التالي بين سنوات الخبرة والراتب الشهري لعينة مكونة من ١٠٠ عاملا أختيروا عشوائيا من إحدى الشركات الصناعية والمطلوب حساب معامل الارتباط (بيرسون) بين سنوات الخبرة والراتب الشهري للعامل .

الراتب (ص) / سنوات الخبرة (س)	١٥٠-	١٨٠-	٢١٠-	٢٤٠-	٢٧٠-٣٠٠	المجموع
٠-	٨	٧				١٥
٥-	٢	٧	١٠	٦		٢٥
١٠-		١	١٩	١٠		٣٠
١٥-			١٣	٢		١٥
٢٠-			٨		٢	١٠
٢٥-٣٠					٥	٥
المجموع	١٠	١٥	٥٠	١٨	٧	١٠٠

الحل :

اذا نظرنا الى شكل توزيـع التكـرارات فـى هـذا الجـدول نجـد أن التكـرارات تتركز حول القطر الرئيسى (من أعلى اليمين الى أدنـى اليسـار) ممـا يـدل بشـكل مبدئـى - علـى وجـود ارتبـاط طـردى بيـن المتغـيرين . وحيـث أن أطـوال الفئـات متساوية لكل من المتغير من س ، ص لذلك سوف نستخدم صيغة الفروق المعدلـة لحساب معامل الارتباط لبيرسون ويتم ذلك على النحو التالى :

أولا :-

بالنسبة للمتغير س : نستخدم التوزيع الهامشى للمتغير س ونحسب منـه المجاميع : (مجـ ح/ س ك س) ، (مجـ ح/ ²س ك س) كما يلى :

فئات	تكرارات ك س	مراكز الفئات (س)	ح س = (س- ١٢,٥)	ح/ س = ح س / ٥	ح/ س ك س	ح/ ²س ك س
-٠	١٥	٢,٥	-١٠	-٢	-٣٠	٦٠
-٥	٢٥	٧,٥	-٥	-١	-٢٥	٢٥
-١٠	٣٠	١٢,٥	صفر	صفر	صفر	صفر
-١٥	١٥	١٧,٥	٥	١	١٥	١٥
-٢٠	١٠	٢٢,٥	١٠	٢	٢٠	٤٠
٢٥-٣٠	٥	٢٧,٥	١٥	٣	١٥	٤٥
المجموع	١٠٠	-	-	-	-٥	١٨٥

من الجدول السابق نجد أن :

مجـ ح/ س ك س = −٥ ، مجـ ح/ ٢ س ك س = ١٨٥ .

ثانياً :

بالنسبة للمتغير ص : نستخدم التوزيع الهامشي للمتغير ص لإيجاد المجاميع مجـ ح/ ص ك ص ، مجـ ح/ ٢ ص ك ص كما يلى :

الفئات	تكرارات ك ص	مراكز الفئات (ص)	ح ص = (ص − ٢٢٥)	ح/ ص = ح ص / ٣٠	ح/ ص ك ص	ح/ ٢ص ك ص
−١٥٠	١٠	١٦٥	−٦٠	−٢	−٢٠	٤٠
−١٨٠	١٥	١٩٥	−٣٠	−١	−١٥	١٥
−٢١٠	٥٠	٢٢٥	صفر	صفر	صفر	صفر
−٢٤٠	١٨	٢٥٥	٣٠	١	١٨	١٨
٣٠٠−٢٧٠	٧	٢٨٥	٦٠	٢	١٤	٢٨
المجموع	١٠٠	−	−	−	−٣	١٠١

من الجدول نجد أن :

مجـ ح/ ص ك ص = −٣ ، مجـ ح/ ٢ص ك ص = ١٠١

ثالثا :

لأيجاد مجـ ح/ س ح/ ص ك س ، ص نعيد كتابة الجدول المزدوج الأصلى مع استبدال فئات المتغير س بالانحرافات المختصرة لها الموجودة فى جدول توزيع س الهامشى أى ح/ س واستبدال فئات المتغير ص بالفروق المعدلة لها الموجوده فى جدول توزيع ص الهامشى أى ح/ ص :

ح ص / ح س	-٢	-١	صفر	١	٢	المجموع
-٢	٨ / ٣٢	٧ / ١٤				٤٦
-١	٢ / ٤	٧ / ٧	١٠ / ٠	٦ / -٦		٥
صفر		١ / ٠	١٩ / ٠	١٠ / ٠		٠
١			١٣ / ٠	٢ / ٢		٢
٢			٨ / ٠		٢ / ٨	٨
٣					٥ / ٣٠	٣٠
المجموع	٣٦	٢١	٠	-٤	٣٨	٩١

من الجدول السابق نجد أن :

مجـ ح/ س ح/ ص ك س ، ص = ٩١ .

وقد تم الحصول على هذه القيمة وذلك بضرب التكرار المشترك (ك س ، ص) فى الفروق المعدلة المقابلة له فى

عمود المتغير س ، أى ح/س ، ثم ضرب الناتج فى الفروق المعدلة المقابلة فى صف المتغير ص ، أى ح/ص . ووضع الناتج داخل القوس فى الركن الأسفل لكل خليه فمثلا فى الخلية (١ ، ١) تم ضرب التكرار المشترك وهو ٨ فى الفرق المعدل (ح/س) المناظر له وهو -٢ فحصلنا على -١٦ والذى تم ضربه فى الفرق المعدل (ح/ص) المناظر له وهو (-٢) لنحصل على ٣٢ ، أى أن حاصل الضرب فى هذه الحالة يكون ٨ × -٢ × -٢ = ٣٢ وهو الرقم المكتوب داخل القوس فى الركن الأسفل من نفس الخلية . وهكذا بالنسبة لباقى خلايا الجدول المزدوج مع اعتبار الخلايا التى ليس بها تكرار مساوية للصفر . ثم يتم تجميع حواصل الضرب هذه لجميع التكرارات لنحصل فى النهاية على مجـ ح/س ح/ص ك س ، ص وهى تساوى - كما سبق أن رأينا ٩١ . ويمكننا الآن ايجاد معامل الارتباط باستخدام صيغة الفروق المعدلة كالآتى :

$$\text{ر} = \frac{\text{مجـ ح}_{\text{س}}\,\text{ح}_{\text{ص}}\,\text{ك}_{\text{س ، ص}} - \dfrac{\text{مجـ ح}_{\text{س}}\,\text{ك}_{\text{س}} \times \text{مجـ ح}_{\text{ص}}\,\text{ك}_{\text{ص}}}{\text{مجـ ك}}}{\sqrt{\left(\text{مجـ ح}^{2}_{\text{س}}\,\text{ك}_{\text{س}} - \dfrac{(\text{مجـ ح}_{\text{س}}\,\text{ك}_{\text{س}})^{2}}{\text{مجـ ك}}\right)\left(\text{مجـ ح}^{2}_{\text{ص}}\,\text{ك}_{\text{ص}} - \dfrac{(\text{مجـ ح}_{\text{ص}}\,\text{ك}_{\text{ص}})^{2}}{\text{مجـ ك}}\right)}}$$

$$= \frac{\text{٩١} - \dfrac{\text{-٥} \times \text{-٣}}{\text{١٠٠}}}{\sqrt{\left[\text{١٨٥} - \dfrac{(\text{-٥})^{2}}{\text{١٠٠}}\right]\left[\text{١٠١} - \dfrac{(\text{-٣})^{2}}{\text{١٠٠}}\right]}} = \text{٠,٦٧}$$

وهذا الناتج يدل على وجود علاقة طردية وقوية نوعا ما بين سنوات الخبرة (س) والراتب الشهرى (ص) .

(٦ - ٢) إرتباط الرتب *Rank Correlation*

قد تكون هناك بعض المتغيرات التى يصعب قياسها رقميا مثل مستويات الذكاء والجمال أو الاداء من أى نوع بالنسبة لمجموعة من المتنافسين يكون قياسها رقميا ممكنا ولكن دقة القياس والنتائج لا تهمنا بدرجة كبيرة ، وفى كلتا الحالتين فاننا نستعيض عن القيم لمفردات المجموعة برتب تدل عليها كأستخدام تقديرات بنجاح الطلاب بدلا من درجاتهم واستخدام صفات دالة على مستوى الأداء وتدرجه من منخفض الى مرتفع والعكس وهكذا . ويمكن فى هذه الحالات ايجاد العلاقة بين المتغيرين على أساس رتب المفردات لديهما بدلا من القيم فاذا

كانت الرتب للمتغيرين تسير في اتجاه واحد كأن تناظر الرتب العالية للمتغير الأول الرتب العالية للمتغير الثاني والعكس كانت العلاقة طردية وعلى العكس اذا كانت الرتب تسير في اتجاه مغاير كانت العلاقة عكسية . ومعامل الارتباط المحسوب لقياس هذا النوع من العلاقات يسمى بمعامل ارتباط الرتب ، وهو أكثر استخداما بالنسبة للعينات الصغيرة من معامل بيرسون الذى يفضل في حالة العينات الكبيرة .

وهناك معاملات لأرتباط الرتب هما معامل سبيرمان ومعامل كندال وسنقدم فيما يلى شرحا لكل منها :

(٦ - ٢ - ١) معامل سبيرمان لأرتباط الرتب :

Spearman Rank Correlation Coefficient:

يعتمد حساب معامل سبيرمان للأرتباط على ترتيب قيم كل من المتغيرين ترتيبا تصاعديا أو تنازليا ففى حالة الترتيب التصاعدى تأخذ أقل قيمة من قيم المتغير الرتبة أو القيمة الاعلى منها الرتبة ، وهكذا والعكس في حالة الترتيب التنازلى حيث تعطى أكبر قيمة من قيم المتغير الرتبة والقيمة الأقل منها الرتبة ، وهكذا وعند تساوى مفردتين من المفردات أو أكثر في القيم التى تأخذها فيحسب متوسط الرتب بافتراض اختلاف القيم وأعطاء كل مفردة منها هذا المتوسط ، فاذا أخذت المفردة الأولى الرتبة أو المفردة الثانية الرتبة ، وإذا تساوت المفردات

الثالثة والرابعة والخامسة فى القيمة فنفترض أولا أن القيم مختلفة ومعنى ذلك أن هذه المفردات تأخذ الرتب ٣ ، ٤ ، ٥ ومتوسطها ١٢ ÷ ٣ = ٤ فتعطى الرتبة ٤ لكل مفرده من هذه المفردات الثلاث ، ثم يكون الترتيب بعد ذلك من الرتبة ٦ وهكذا .

وبعد ترتيب المتغيرين نوجد الفرق بين أزواج الرتب للمتغيرين ويرمز لهذا الفرق بالرمز " ف " ثم نربع هذه الفروق أى نحصل على ($ف^2$) ونوجد مجموعها ونطبق فى قانون معامل ارتباط سبيرمان التالى :

$$ر = ١ - \frac{٦ \text{ مج } ف^2}{ن (ن^2 - ١)}$$

ويلاحظ أن مجموع الفروق بين أزواج الرتب = صفر دائما

كما أن مجموع الرتب متغير من المتغيرين = $\frac{ن (ن + ١)}{٢}$

وتنحصر قيمة معامل سبيرمان لأرتباط الرتب بين (١ ، -١) كما هو الحال فى معامل بيرسون للأرتباط .

<u>مثال (٥) :</u>

الاتى ببيان بدرجات ٦ طلاب فى مادتى الأحصاء والرياضة والمطلوب حساب معامل سبيرمان لأرتباط الرتب .

درجات الأحصاء	١٨	١٤	١٦	١٠	٧	٩
درجات الرياضة	١٣	١٠	١٨	٥	٥	٩

الحل

درجات الأحصاء (س)	درجات الرياضة (ص)	رتب الاحصاء	رتب الرياضة	فروق الرتب (ف)	مربعات الفروق (ف٢)
١٨	١٣	١	٢	-١	١
١٤	١٠	٣	٣	صفر	صفر
١٦	١٨	٢	١	١	١
١٠	٥	٤	٥,٥	-١,٥	٢,٢٥
٧	٥	٦	٥,٥	٠,٥	٠,٢٥
٩	٩	٥	٤	١	١
مجـ	-	٢١	٢١	صفر	٥,٥

∴ ن (ن + ١) / ٢ = ٦ × ٧ / ٢ = ٢١ = مجموع الرتب لكل من س ، ص

∴ ر = ١ - ٦ × ٥,٥ / ٦ (٣٦ - ١)

= ١ - ٣٣ / ٢١٠

= ١ - ٠,١٥٧

= ٠,٨٤٣

مما يدل على وجود ارتباط طردى قوى بين درجات الطلاب فى كل من مادتى الاحصاء والرياضة .

مثال (٦) :

فيما يلى تقديرات ثمانية من الطلاب فى امتحان مادتى الرياضيات والاحصاء

والمطلوب : حساب معامل ارتباط الرتب بين تقديرات المادتين .

تقديرات الرياضيات	ضعيف	مقبول	ممتاز	مقبول	ضعيف	جيد جدا	جيد	مقبول
تقديرات الاحصاء	مقبول	جيد	جيد جدا	جيد	مقبول	مقبول	ممتاز	ضعيف

الحل :

أولا : نرتب تقديرات مادة الرياضيات تصاعديا بحيث تكون ضعيف ، مقبول ، جيد ، جيد جدا ، ممتاز وكذلك الأمر بالنسبة لتقديرات مادة الاحصاء بحيث نحصل على الجدول التالى :

تقديرات الرياضيات (س)	تقديرات الاحصاء (ص)	رتب س	رتب ص	فروق الرتب (ف)	مربعات الفروق (ف٢)
ضعيف	مقبول	١,٥	٢,٥	١	١
مقبول	جيد	٤	٥	١	١
ممتاز	جيد جدا	٨	٧	١	١
مقبول	جيد	٤	٥	١	١
ضعيف	مقبول	١,٥	٢,٥	١	١
جيد جدا	جيد	٧	٥	٢	٤
جيد	ممتاز	٦	٨	٢	٤
مقبول	ضعيف	٤	١	٣	٩
-	-	-	-	-	مجـ ف٢ = ٢٢

$$\text{ر} = ١ - \frac{٦ \ \text{مجـ ف}^{٢}}{\text{ن} \left(\text{ن}^{٢} - ١ \right)}$$

$$= ١ - \frac{٦ \times ٢٢}{٨ \left(٦٤ - ١ \right)} = ٠,٧٤$$

مما يدل على وجود ارتباط طردى قوى بين تقديرات الرياضيات وتقديرات الاحصاء .

١-٢-٢ معامل كندال لأرتباط الرتب :

Kendall Rank Correlation Coefficient

معامل كندال لأرتباط الرتب والذي نظيره معامل سبيرمان من الناحيتين

النظرية والعملية [illegible] عن العلاقة بين المتغيرين [illegible]

[illegible] الرتب المتناظرة [illegible] الرتب ثم اعادة ترتيب [illegible]

[illegible] (ص مثلا) التي يتضمن الترتيب من الأعداد الطبيعية أي نبدأ بالرتبة ١ ثم ٢ ثم ٣ وهكذا ومع كل رتبة نكتب الرتبة المناظرة لها للمتغير الآخر ثم نبدأ في [illegible] في رتب المتغير الآخر (ص) فنبدأ بالرتبة الأولى وننظر الى الرتبة المناظرة لها في رتب (س) ونطرح [illegible] التي على يمينها من عدد الرتب التي على يسارها ونحذف هذا الزوج من الرتب ثم ننتقل الى الرتبة الثانية من رتب (ص) وننظر الى الرتبة المناظرة لها في رتب (س) ونطرح كذلك عدد الرتب التي على يمينها من عدد الرتب التي على يسارها ونحذف بعد ذلك هذا الزوج من الرتب أيضا ثم نكرر هذه العملية حتى ننتهي من جميع ازواج الرتب التي على اليسار وتلك التي على اليمين ونطبق في القانون التالي :-

$$\text{معامل كندال لارتباط الرتب} = \frac{\text{٢ مجـ ف}}{\text{ن (ن − ١)}}$$

حيث : ف تمثل الفروق بين عدد الرتب على الجانبين

. ن هي عدد أزواج الرتب .

مثال (٧) :

الاتى بيان بدرجات عشرة من الطلاب فى مادتى الاحصاء والرياضة والمطلوب حساب معامل كندال لارتباط الرتب بين درجات المادتين .

درجات الأحصاء	٧٨	٨١	٨٥	٩٠	٨٧	٧٥	٦٤	٦٤	٦٠	٨٠
درجات الرياضة	٨٠	٩٤	٧٤	٧٠	٧٥	٦٥	٥٩	٦٢	٨٦	٩٢

الحل

١- نقوم بترتيب درجات كل من المادتين تصاعديا أو تنازليا .

رتب الأحصاء	٦	٤	٣	١	٢	٧	٨,٥	٨,٥	١٠	٥
رتب الرياضة	٤	١	٦	٧	٥	٨	١٠	٩	[illegible]	٢

٢- نقوم باعادة ترتيب أحد المتغيرين حسب الأعداد الطبيعية وتكتب معها رتب المتغير الأخر المناظرة ، فاذا قمنا باعادة كتابة رتب الرياضة (ص) حسب الاعداد الطبيعية

رتب الأحصاء (س)	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
رتب الرياضة (ص)	٤	٥	١٠	٦	٢	٣	١	٧	٨,٥	٨,٥

٣- نبحث فى رتب (ص) ونبدأ بالرتبة ١ ويناظرها الرتبة ٧ فى رتب المتغير س ونجد عدد الرتب على يسارها (الأعلى منها) = ٣

وعدد الرتب التى على يمينها (الأقل منها) = ٦ فنقوم بطرح الثانية من الأولى أى ٣ - ٦ = -٣ ونحذف هذا الزوج من الرتب ، ثم ننظر الى الرتبة الثانية فى رتب (ص) ويناظرها الرتبة ٥ من رتب (س) ونجد أن عدد الرتب الأعلى منها = عدد الرتب الأقل منها = ٤ اذن فالفرق = صفر ثم نحذف هذا الزوج ايضا

وهكذا نكرر هذه العملية حتى ننتهى من جميع أزواج الرتب فنحصل على الفرق -٣ ، صفر ، -١ ، ٦ ، ٥ ، ٢ ، ١ ، صفر ، -١ .

٤- نحسب مجموع الفروق أى نوجد مجـ ف فنجدها = ٩ نطبق فى العلاقة السابقة لمعامل كندال على النحو التالى :

$$ر = \frac{٢ \text{ مجـ ف}}{ن (ن - ١)} = \frac{٢ \times ٩}{١٠ \times ٩} = \frac{١٨}{٩٠} = ٠,٢$$

وهذا يدل على أن العلاقة بين درجات المادتين موجبة ولكنها ضعيفة .

حينما تكون احدى الظاهرتين المراد بحث العلاقة بينهما أو كلتاهمـا لا يتـم التعبير عنها رقميا بل تنقسم الظاهرة الى عدة أقسام يعبر عن كل منها بصفـة أو خاصية معينة ويحدد عدد المفردات التى تنتمى لكل صفة أو خاصيـة (تكـرار كـل صفـة) فـان هنـاك مقـاييس خاصـة للأرتبـاط بيـن هذه الظواهـر غـير المقـاييس المستخدمة فى قياس الارتباط بين الظواهر الكمية (معامل بيرسون) أو الترتيبيـة (معـاملا سبيرمان وكنـدال) . ومـن أمثلـة العلاقـات الوصفيـة العلاقـة بيـن تفكـك الروابط الاسرية والانحلال الخلقى والعلاقة بين التدخيـن وسـرطان الرئـة والعلاقـة بين مستوى الذكاء وسوءالتغذية وغيرها من الامثلة .

وتبوب الظاهرتان فى هذه الحالات فـى صـورة جـدول مـزدوج فـى الصـف الأول تظهر أقسام احدى الظاهرتين وفى العمود الأول تظهر أقسام الظاهرة الثانيـة وتتكون خلايا الجدول من التكرارات التى يمثل كل منها عدد المفردات التـى تنتمـى لقسم معين من الظاهرة الأولى وقسم أخر من الظاهرة الثانية ويعتمد قياس العلاقة بين الظاهرتين على هذه التكرارات دون أى شىء أخر . وهناك مقياسان معروفـان للأرتبـاط بيـن الصفـات همـا معـامل الاقـتران ومعـامل التوافـق . وفيمـا يلـى شـرح لكيفية حساب كل منهما :

(٦-٣-١) معامل الاقتران

يستخدم ذلك المقياس اذا كانت كل من الظاهرتين المراد بحث العلاقة بينهما تنقسم الى صفتين أي خاصيتين بحيث يمكن وضع البيانات الخاصة بهما في صورة جدول يحتوى على أربع خلايا (جدول (٢ × ٢)) وفيه ترتب صفات احدى الظاهرتين أفقيا وصفت الاخرى رأسيا ويأخذ الصورة العامة التالية :

الظاهرة الأولى / الظاهرة الثانية	خاصية (١)	خاصية (٢)
خاصية (١)	أ	ب
خاصية (٢)	جـ	د

حيث أ ، ب ، جـ ، د هي تكرارات الخواص الاربع ويحسب معامل الأقتران في هذه الحالة من العلاقة الاتية :

$$\text{معامل الاقتران} = \frac{\text{أ} \times \text{د} - \text{ب} \times \text{جـ}}{\text{أ} \times \text{د} + \text{ب} \times \text{جـ}}$$

وتنحصر قيمة هذا المعامل بين +١ ، -١ .

فاذا كانت قيمة أى تكرار على القطر الرئيسى (أى قيمة أ ، د) مساوية للصفر فان معامل الاقتران سوف تصبح قيمته -١ وهذا يعنى أرتباط تام وسالب بين الظاهرتين ، واذا كانت قيمة أى تكرار على القطر المعاكس (أى قيمة ب أو جـ) مساوية للصفر فان قيمة المعامل سوف تصبح +١ وهذا يعنى أرتباط تام

موجب أما اذا كان حاصل ضرب التكرارات على القطر الرئيسى يساوى حاصل ضرب التكرارات على القطر المعاكس أى أن أ د = ب جـ فان قيمة معامل الاقتران تصبح مساوية للصفر مما يعنى عدم وجود علاقة بين الظاهرتين . وكلما اقتربت قيمة المعامل من الواحد الصحيح كلما دل ذلك على قوة العلاقة بينهما وكلما اقتربت القيمة من الصفر كلما دل ذلك على ضعف العلاقة .

__مثال (٨) :__

اذا كنا بصدد بحث العلاقة بين لون نوع معين من الزهور ورائحته وكانت لدينا ١٠٠ زهرة منها قسمت حسب اللون والرائحة كما فى الجدول التالى فما نوع العلاقة بينهما ؟

اللون \ الرائحة	رائحة زكية	بدون رائحة
أحمر	٣٠	١٠
أصفر	٣٥	٢٥

الحل أ = ٣٠ ، ب = ١٠ ، جـ = ٣٥ ، د = ٢٥

$$\text{معامل الاقتران} = \frac{٣٠ \times ٢٥ - ٣٥ \times ١٠}{٣٠ \times ٢٥ + ٣٥ \times ١٠} = \frac{٤٠٠}{١١٠٠} = ٠,٣٦$$

معنى ذلك أن العلاقة بين لون الزهر ورائحته علاقة طردية ولكنها ضعيفة.

(٦-٣-٢) معامل التوافق :

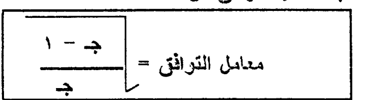
اذا كانت احدى الظاهرتين المراد بحث العلاقة بينهما أو كلتاهما مقسمة الى أكثر من صفتين أو خاصيتين بحيث يكون عدد خلايا الجدول المحتوية على التكرارات أكثر من ٤ فانه لا يمكننا حساب معامل الاقتران ونحسب بدلا منه معامل التوافق والذى يتكون من الاجراءات التالية :

(١) تربيع التكرار الموجود بكل خليه من خلايا الجدول .

(٢) قسمة مربع التكرار بالخلية على حاصل ضرب مجموع التكرارات بالصف الذى توجد به الخليه فى مجموع التكرارات بالعمود الذى توجد به الخلية.

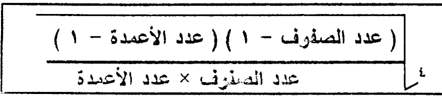
(٣) تكرار الخطوتين السابقتين لكل خلايا الجدول وايجاد مجموع خوارج القسمة ولنرمز لهذا المجموع بالرمز جـ .

(٤) حساب معامل التوافق من العلاقة :



$$\text{معامل التوافق} = \sqrt{\frac{\text{جـ} - 1}{\text{جـ}}}$$

وتنحصر قيمة معامل التوافق بين الصفر وبين الحد الأعلى له والذى يساوى القيمة التالية :



$$\sqrt[4]{\frac{(\text{عدد الصفوف} - 1)(\text{عدد الأعمدة} - 1)}{\text{عدد الصفوف} \times \text{عدد الأعمدة}}}$$

وكلما اقتربت قيمة المعامل من الحد الأعلى كلما دل ذلك على قوة العلاقة.

وكلما اقتربت قيمته من الصفر كلما دل ذلك على ضعف العلاقة .

ويلاحظ أن معامل التوافق لا يحدد اتجاه العلاقة أى موجبة أو سالبة .

مثال (٩) :

الجدول التالى يبين توزيع ٣٠٠ شخص حسب المستوى التعليمى والتدخين والمطلوب قياس العلاقة بينهما .

المستوى التعليمى \ التدخين	يدخن	لا يدخن	المجموع
متعلم	٧٥	١٥	٩٠
يقرأ ويكتب	٩٠	٧٠	١٦٠
أمى	١٥	٤٥	٦٠
المجموع	١٨٠	١٢٠	٣٠٠

الحل :

نوجد أولا قيمة (جـ) على النحو التالى :

$$\text{جـ} = \frac{٧٥ \times ٧٥}{٩٠ \times ١٨٠} + \frac{٩٠ \times ٩٠}{١٥٠ \times ١٨٠} + \frac{١٥ \times ١٥}{٦٠ \times ١٨٠}$$

$$+ \frac{١٥ \times ١٥}{٩٠ \times ١٢٠} + \frac{٦٠ \times ٦٠}{١٢٠ \times ١٢٠} + \frac{٤٥ \times ٤٥}{٦٠ \times ١٢٠}$$

$$= ٠,٣٤٧ + ٠,٣٠٠ + ٠,٠٢١ + ٠,٠٢١ + ٠,٢٥٠ + ٠,٢٨١$$

$$= ١,٢٢٠$$

$$\text{معامل التوافق} = \sqrt{\frac{١,١٧٠ - ١}{١,١٧٠}} = ٠,٣٨$$

أما الحد الأعلى لمعامل التوافق هو :

$$\sqrt[٤]{\frac{(٣-١)(٢-١)}{٣ \times ٢}}$$

$$= \sqrt[٤]{\frac{١}{٣}} = ٠,٧٦$$

وتكون النسبة بين قيمة المعامل وحده الأعلى هى :

$$= \frac{٠,٣٨}{٠,٧٦} = ٠,٥$$

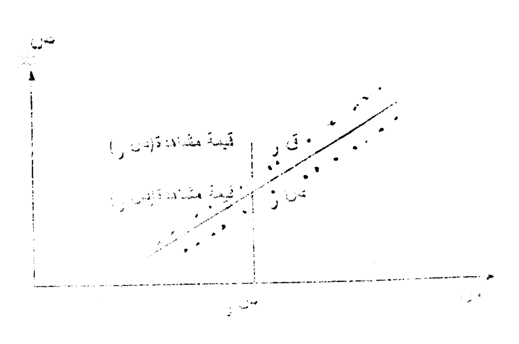
أى أنه توجد علاقة بين المستوى التعليمى للشخص والتدخين وهذه العلاقة بين درجة متوسطة .

الباب السابع

الانحدار الخطي

Linear Regression

تكلمنا في الباب السادس عن الارتباط ومعناه وكيفية قياسه في الحالات المختلفة ، وعرفنا أن مقاييس الارتباط تمكننا من معرفة درجة العلاقة بين ظاهرتين أو متغيرين ، وعرفنا أنه إذا كان الارتباط تاما بين ظاهرتين فإن قيمهما تقع على خط مستقيم في شكل الانتشار وإذا كان الارتباط منعدما فإن هذه القيم تكون مبعثرة بحيث لا يوجد رابط بينها . وكلما كان الارتباط قويا كلما انحدرت أو قربت قيم الظاهرتين من خط مستقيم يمثل العلاقة بينهما ويسمى هذا المستقيم بخط الانحدار ويفيدنا هذا الخط في عملية التنبؤ بقيم المتغير التابع (ص) بمعلومية قيم المتغير المستقل أو المتغيرات المستقلة (س).



فعملية التنبؤ بالقيم المستقبلية للظواهر غاية فى الاهمية خصوصا فى مجال التخطيط فى الاقتصاد والادارة والسياحة والتعليم والصحة والقوى العاملة ..... الخ . فتخصيص الميزانية لخطة خمسية قادمة يتطلب التنبؤ بعدد السكان المتوقع فى المراحل التعليمية المختلفة ، وأعداد الخريجين وقوة العمل وحجم المبانى والالآت والأجهزة المطلوبة لتنفيذ المستهدف من الخطة ، وتلعب معادلات الانحدار الدور الرئيسى فى الوصول الى هذه التنبؤات كما سنرى فيما بعد .

(٧-١) الانحدار الخطى بين المتغيرين (س ، ص)

[فى حالة البيانات الغير مبوبة] :

نفرض أنه لدينا متغيرين (س ، ص) بينهما علاقة خطية ولهما القيم الآتية (س١ ، ص١) ، (س٢ ، ص٢) ، .... (س ن ، ص ن) وأردنا الحصول على معادلة أحسن خط مستقيم يمثل هذه البيانات ، لوجدنا أنه الخط الذى يمر بأغلب النقاط ويمر بتوازن بين بقية النقاط أو بمعنى آخر هو ذلك الخط الذى يكون **التشتت حوله أقل ما يمكن . توجد عدة طرق لأيجاد ذلك الخط منها طريقة** المربعات الصغرى *Least Squares* وأساس هذه الطريقة أختيار الخط الذى يطابق النقاط أحسن مطابقة هو الخط الذى يكون مجموع مربعات انحرافات النقاط عنه أصغر ما يمكن (أى نهاية صغرى) . [illegible] فى الانحدار [illegible] الأول منهما يكون (ص) متغير [illegible] بينما يكون (س) متغير تابع .

(٧-١-١) خط انحدار ص على س :

أذا كان (س) متغير مستقل ، (ص) متغير تابع فان معادلة شكل الأنتشار ستكون كالآتى :

$$\text{ص} = \text{أ}\,\text{س} + \text{ب} + \text{ق}$$

ويكون مجموع مربعات الأخطار هو .

$$\text{مجـ ق}^{٢} = \text{مجـ (ص} - \text{أ س} - \text{ب)}^{٢}$$

ولجعل هذه الأخطار نهاية صغرى فاننا نفاضل جزئيا مره بالنسبة الى (أ) ومرة آخرى بالنسبة الى (ب) كالآتى :

$$\frac{\delta\ \text{مجـ ق}^{٢}}{\delta\ \text{أ}} = -٢\ \text{مجـ س (ص} - \text{أ س} - \text{ب)}$$

$$\frac{\delta\ \text{مجـ ق}^{٢}}{\delta\ \text{ب}} = -٢\ \text{مجـ (ص} - \text{أ س} - \text{ب)}$$

وبمساواة التفاضلات الجزئية السابقة بالصفر نحصل على المعادلتين الأتيتين :

$$\text{مجـ س ص} = \hat{\text{أ}}\ \text{مجـ س}^{٢} + \hat{\text{ب}}\ \text{مجـ س}$$

$$\text{مجـ ص} = \hat{\text{أ}}\ \text{مجـ س} + \hat{\text{ب}}\ \text{ن}$$

ويطلق عليهما أسم المعادلات المعتادة *Normal Equations* ثم وضع العلاقة (^) فرق أ ، ب لانه لا توجد سوى قيمة واحده لـ أ ، ب تجعل متوسط مربعات أنحرافات القيم حول الخط نهاية صغرى ورمزنا بهم بالرمز أ ، ب . وبحل المعادلتين السابقتين نحصل على قيمة أ ، ب كالاتى :

$$\hat{أ} = \frac{\text{مجـ} (س - \bar{س})(ص - \bar{ص})}{\text{مجـ} (س - \bar{س})^{٢}}$$

$$= \frac{\text{مجـ س ص} - \frac{\text{مجـ س} \times \text{مجـ ص}}{ن}}{\text{مجـ س}^{٢} - \frac{(\text{مجـ س})^{٢}}{ن}}$$

$$\hat{ب} = \bar{ص} - أ \bar{س}$$

ويسمى (أ) بمعامل انحدار *Coefficient of Regression* ص على س

ويرمز له أحيانا بالرمز $أ_{ص / س}$. وبذلك تكون معادلة أحسن خط أنحدار هى :

$$\hat{ص} = \hat{أ} س + \hat{ب}$$

مثال (١) :

من البيانات التالية أحسب معادلة انحدار ص/ س حيث :

س	١	٢	٣	٤	٥
ص	٣	٦	٩	١٢	١٥

الحل : لايجاد معادلة انحدار ص / س نكون الجدول التالي :

س	ص	س × ص	س٢	ص٢
١	٣	٣	١	٩
٢	٦	١٢	٤	٣٦
٣	٩	٢٧	٩	٨١
٤	١٢	٤٨	١٦	١٤٤
٥	١٥	٧٥	٢٥	٢٢٥
مجـ =١٥	٤٥	١٦٥	٥٥	٤٩٥

نفرض أن أحسن خط انحدار يمثل هذه البيانات هو :

$$\hat{\text{ص}} = \hat{\text{أ}}\ \text{س} + \hat{\text{ب}}$$

$$\text{حيث } \hat{\text{أ}} = \frac{\text{مجـ س ص} - \dfrac{\text{مجـ س} \times \text{مجـ ص}}{\text{ن}}}{\text{مجـ س}^{٢} - \dfrac{(\text{مجـ س})^{٢}}{\text{ن}}}$$

$$= \frac{165 - \frac{15 \times 45}{5}}{55 - \frac{15 \times 15}{5}} = \frac{30}{10} = 3$$

وحيث أن متوسط س ومتوسط ص هما :

$$\bar{س} = \frac{15}{5} = 3 \quad , \quad \bar{ص} = \frac{45}{5} = 9$$

فان الجزء الثابت ($\hat{ب}$) يمكن الحصول عليه كالآتى :

$$\hat{ب} = \bar{ص} - \hat{أ}\,\bar{س}$$

$$= 9 - 3 \times 3 = \text{صفر}$$

وبذلك تكون معادلة خط انحدار ص / س هى :

$$\hat{ص} = 3\,س$$

<u>(٧-١-٢) خط انحدار س على ص :</u>

بافتراض (ص) فى هذه الحالة هو المتغير المستقل وأن (س) هـو المتغير التابع وبالتالى فإن معادلة أحسن خط انحدار هى :

$$\hat{س} = \hat{أ}_1\,ص + \hat{ب}_1$$

وباستخدام اسلوب المربعات الصغرى السابقة يمكننا الحصول على $\hat{أ}_١$ ،

$\hat{ب}_١$ كما يلى :

$$\hat{أ}_١ = \frac{\text{مجـ}(س - \bar{س})(ص - \bar{ص})}{\text{مجـ}(ص - \bar{ص})^٢}$$

أو

$$\hat{أ}_١ = \frac{\text{مجـ س ص} - \frac{\text{مجـ س} \times \text{مجـ ص}}{ن}}{\text{مجـ ص}^٢ - \frac{(\text{مجـ ص})^٢}{ن}}$$

$$، \hat{ب}_١ = \bar{س} - \hat{أ}_١ \bar{ص}$$

حيث يطلق على $\hat{أ}_١$ معامل انحدار س / ص أو يرمز به بالرمز $\hat{أ}_١$ س / ص

مثال (٢) :

من بيانات المثال السابق أوجد معادلة انحدار س / ص

الحل

باستخدام طريقة المربعات الصغرى يمكننا الحصول على $\hat{أ}_١$ ، $\hat{ب}_١$ كما يلى:

$$\hat{أ}_١ = \frac{\text{مجـ س ص} - \frac{\text{مجـ س} \times \text{مجـ ص}}{ن}}{\text{مجـ ص}^٢ - \frac{(\text{مجـ ص})^٢}{ن}}$$

$$= \frac{١٦٥ - \frac{١٥ \times ٤٥}{٥}}{٤٩٥ - \frac{٤٥ \times ٤٥}{٥}} = \frac{١٦٥ - ١٣٥}{٤٩٥ - ٤٠٥} = \frac{٣٠}{٩٠} = ٠,٣٣٣$$

$$\hat{ب}_{١} = \bar{س} - \hat{أ}_{١} \bar{ص}$$

$$= ٣ - ٠,٣٣٣ \ (٩)$$

$$= ٣ - ٢,٩٧٧ = ٠,٠٠٣$$

وبذلك تكون معادلة خط انحدار س / ص هى :

$$\hat{س} = ٠,٣٣٣ \ ص + ٠,٠٠٣$$

<u>ملاحظات على معاملات الانحدار (ص / س) ، (س / ص) :</u>

١ - نلاحظ من المثالين السابقين أنه على الرغم من الاختلاف قيمة كل من $\hat{أ}$ ص / س ، $\hat{أ}_{١}$ س / ص الا أنهما لهما نفس الاشارة ويمكن تفسير ذلك كالاتى:

حيث أن :

$$\hat{أ} \ ص / س = \frac{تغاير \ (س \ ، ص)}{تباين \ س} \ ، \ \hat{أ}_{١} \ س / ص = \frac{تغاير \ (س \ ، ص)}{تباين \ ص}$$

ويرجع ثبات الاشارة الى أن البسط فى المقدارين (تغاير س ، ص) هو الذى يحدد اشارة المقدار .

٢ - يتقاطع كل من خطى انحدار ص / س ، س / ص اذا رسما على نفس شكل الانتشار ويتحقق ذلك عند النقطة ($\bar{س}$ ، $\bar{ص}$)

٣ - يمكننا الحصول على مربع معامل الارتباط البسيط عن طريق ضرب معامل انحدار ص/ س فى معامل انحدار س / ص أى أن :

$$ر^2 = \hat{أ} \times \hat{أ}_1$$

وبالتالى فإن

$$\text{معامل الارتباط (ر)} = \sqrt{\hat{أ} \times \hat{أ}_1}$$

مثال (٣)

من نتائج المثالين (١) ، (٢) أوجد معامل الارتباط (ر) متحققا من النتيجة باستخدام بيانات المثال (١) .

الحل

حيث أن $\hat{أ} = ٣$ ، $\hat{أ}_1 = .٣٣٣$

$\therefore$ ر $= \sqrt{٣ \times ٠,٣٤} = ١$ تقريبا

ولذا يمكننا إيجاد قيمة (ر) باستخدام القانون مباشرة على النحو التالي:

$$= \frac{\text{مجـ س ص} - \frac{\text{مجـ س} \times \text{مجـ ص}}{\text{ن}}}{\sqrt{\left(\text{مجـ س}^{2} - \frac{(\text{مجـ س})^{2}}{\text{ن}}\right)\left(\text{مجـ ص}^{2} - \frac{(\text{مجـ ص})^{2}}{\text{ن}}\right)}}$$

$$= \frac{١٦٥ - \frac{١٥ \times ٤٥}{٥}}{\sqrt{\left(٥٥ - \frac{(١٥)^{٢}}{٥}\right)\left(٤٩٥ - \frac{(٤٥)^{٢}}{٥}\right)}} = ١$$

وهى نفس النتيجة السابقة باستخدام العلاقة بين معاملى انحدار ص / س ، س / ص .

ومما يجب ملاحظته أنه يمكننا حساب معامل الارتباط (ر) أيضا باستخدام العلاقات التالية :

$$\text{ر} = \text{أ}_{٢} \times \frac{\text{ع س}}{\text{ع ص}} \qquad \text{أو} \qquad \text{ر} = \text{أ}_{١} \times \frac{\text{ع ص}}{\text{ع س}}$$

حيث :

ع س = الانحراف المعياري للمتغير (س) $= \sqrt{\frac{\text{مجـ}(\text{س} - \bar{\text{س}})^{2}}{\text{ن}}}$

$$ع_ص = \text{الانحراف الميعارى للمتغير ص} = \sqrt{\frac{مجـ (ص - \bar{ص})^2}{ن}}$$

كذلك وباجراء بعض العمليات الجبرية فانه يمكننا ايجاد معاملات الانحدار

ص / س ، س / ص من خلال العلاقتين السابقتين كما يلى :

$$\text{معامل انحدار ص / س } (أ) = ر \frac{ع_ص}{ع_س}$$

$$\text{معامل انحدار س / ص } (\hat{أ_١}) = ر \frac{ع_س}{ع_ص}$$

مما تقدم يمكننا التوصل الى صيغ اخرى لمعادلات الانحدار وذلك بدلالة كل من معامل الارتباط ومتوسطات المتغيرات ($\bar{س}$ ، $\bar{ص}$) وأيضا الانحرافات المعيارية لكل منهما على النحو التالى :

أولا : معادلة انحدار ص / س :

$$\left(\frac{\hat{ص} - \bar{ص}}{ع_ص}\right) = ر \left(\frac{س - \bar{س}}{ع_س}\right)$$

ثانيا معادلة انحدار س / ص :

$$\left(\frac{\hat{س} - \bar{س}}{ع_س}\right) = ر \left(\frac{ص - \bar{ص}}{ع_س}\right)$$

مثال (٤) :

مستخدما بيانات المثال (١) أوجد ما يلي :

- معادلات الانحدار ص / س ، س / ص

- معامل الارتباط بين المتغيرين س ، ص .

- معادلات انحدار ص / س ، س / ص .

الحل :

من بيانات المثال (١) نجد أن :

$$\bar{\text{س}} = \frac{\text{مجـ س}}{\text{ن}} = \frac{١٥}{٥} = ٣$$

$$\bar{\text{ص}} = \frac{\text{مجـ ص}}{\text{ن}} = \frac{٤٥}{٥} = ٩$$

$$\text{ع}_{\text{س}} = \sqrt{\frac{\text{مجـ س}^٢}{\text{ن}} - (\bar{\text{س}})^٢} = \sqrt{\frac{٥٥}{٥} - (٣)^٢} = \sqrt{١١ - ٩} = \sqrt{٢}$$

$$\text{ع}_{\text{ص}} = \sqrt{\frac{\text{مجـ ص}^٢}{\text{ن}} - (\bar{\text{ص}})^٢} = \sqrt{\frac{٤٩٥}{٥} - (٩)^٢} = \sqrt{٩٩ - ٨١} = \sqrt{١٨}$$

$\because$ ر = ١

$\therefore$ معامل انحدار ص / س (أ) = ر . $\frac{\text{ع ص}}{\text{ع س}} = ١ . \frac{\sqrt{١٨}}{\sqrt{٢}} = ٣$

$$\therefore \text{معامل انحدار ص / س (أ)} = \text{ر} \frac{\text{ع س}}{\text{ع ص}} = ١. \frac{\sqrt{٢}}{\sqrt{١٨}} = \frac{١}{٣}$$

ويمكن استخدام الصيغة الرياضية الاتية فى ايجاد معامل الارتباط (ر) بدلالة مثلا معامل انحدار ص / س والانحرافات المعيارية للمتغيرين س ، ص كما يلى :

$$\text{ر} = \hat{\text{أ}} \times \frac{\text{ع س}}{\text{ع ص}}$$

$$= ٣ \times \frac{\sqrt{٢}}{\sqrt{١٨}} = ٣ \times \sqrt{\frac{١}{٩}} = ٣ \times \frac{١}{٣} = ١$$

وهى نفس النتيجة التى امكن التوصل اليها سابقا .

اما بخصوص معادلات انحدار ص / س ، س / ص فيمكن ايجادها على النحو التالى :

أولا : معادلة انحدار ص / س :

$$\left(\frac{\hat{\text{ص}} - \bar{\text{ص}}}{\text{ع ص}}\right) = \text{ر} \left(\frac{\text{س} - \bar{\text{س}}}{\text{ع س}}\right)$$

$$\therefore \left(\frac{\hat{\text{ص}} - ٩}{\sqrt{١٨}}\right) = ١ \left(\frac{\text{س} - ٣}{\sqrt{٢}}\right)$$

بضرب الطرفين في الوسطين

$$\therefore (\hat{\text{ص}} - ٩)\sqrt{٢} = (\text{س} - ٣)\sqrt{١٨}$$

وبقسمة الطرفين على $\sqrt{٢}$

$$\frac{(\hat{\text{ص}} - ٩)\sqrt{٢}}{\sqrt{٢}} = \frac{(\text{س} - ٣)\sqrt{١٨}}{\sqrt{٢}}$$

$$\therefore \hat{\text{ص}} - ٩ = ٣\,\text{س} - ٩$$

$$\therefore \hat{\text{ص}} = ٣\,\text{س}$$

وبالمثل يمكننا ايجاد معادلة انحدار س / ص على النحو التالى :

$$\frac{\hat{\text{س}} - \bar{\text{س}}}{\text{ع س}} = \text{ر} \,.\, \frac{\text{ص} - \bar{\text{ص}}}{\text{ع ص}}$$

$$\left(\frac{\hat{\text{س}} - ٣}{\sqrt{٢}}\right) = ١\left(\frac{\text{ص} - ٩}{\sqrt{١٨}}\right)$$

بضرب الطرفين في الوسطين

$$\therefore (\hat{\text{س}} - ٣)\sqrt{١٨} = (\text{ص} - ٩)\sqrt{٢}$$

وبقسمة الطرفين على $\sqrt{18}$ نحصل على

$$\hat{س} = \frac{1}{3} ص + 0.003$$

ملاحظات حول معامل الانحدار :

١- معامل الانحدار لا يتأثر بعملية الجمع أو الطرح

حيث أن :

$$\text{معامل الانحدار } (ص / س) = \frac{\text{تغا } (س / ص)}{\text{تبا } (س)}$$

أو

$$\text{معامل الانحدار } (س / ص) = \frac{\text{تغا } (س ، ص)}{\text{تبا } (ص)}$$

وقد سبق أن أوضحنا فى الباب السابق أن كلا من التغاير أو التباين لا يتأثرا بعمليتى الجمع أو الطرح وبالتالى فان معاملات الانحدار سواء (ص / س) أو (س / ص) لا يتأثرا أيضا بالجمع أو الطرح .

٢- معامل الانحدار يتأثر بعمليات الضرب والقسمة .

حيث أنه اذا ضربنا قيم (س) فى الثابت (أ) وقيم (ص) فى الثابت (ب) فان :

تغا (أ س ، ب ص) = أ ب تغا (س ، ص)

، تبا (أ س) = أ^2 تبا (س)

، تبا (ب ص) = ب^2 تبا (ص)

وبذلك فان :

$$\hat{\text{أ ب ص / أ س}} = \frac{\text{تغا (ب ص ، أ س)}}{\text{تبا (أ س)}}$$

$$= \frac{\text{أ ب تغا (ص ، س)}}{\text{أ}^2 \text{ تبا (س)}}$$

$$= \frac{\text{ب}}{\text{أ}} \; \frac{\text{تغا (ص ، س)}}{\text{تبا (س)}}$$

$$= \frac{\text{ب}}{\text{أ}} \; \hat{\text{أ ص / س}}$$

وبذلك نستطيع استنتاج أنه اذا تم ضرب قيم (س) فى نفس المقدار الثابت الأول ثم ضرب المتغير (ص) فيه أي أن (أ = ب) فإن معامل الانحدار لا يتأثر في هذه الحالة .

٣- على العكس من معامل الارتباط فان معامل الانجدار يمكن أن يأخذ قيما أكبر من + ١ أو قيما أصغر من - ١ لأنه عبارة عن معدل تغير ويمكن أن يأخذ أي قيمة ، واذا كان معامل الانجدار مساويا للصفر فهذا يدل على عدم وجود علاقة انحدار بين المتغيرين س ، ص .

الجدول التالى يبين دخل ٨ أسر ومقدار ما تنفقه من هذا الدخل

الدخل بالجنيه	٦٤	٥٢	٨٤	٦٤	٧٦	٥٦	٦٨	٥٤
الانفاق بالجنيه	٥٢	٤٠	٦٠	٥٢	٦٠	٤٢	٥٠	٥٢

المطلوب ايجاد :

١- معامل الارتباط بيرسون .

٢- خط انحدار الانفاق على الدخل (ص / س) .

٣- خط انحدار الدخل على الانفاق (س / ص) .

٤- قدر انفاق الاسرة التى يبلغ دخلها ٧٠٠ جنيه .

٥- قدر دخل الاسرة التى تنفق ٤٨٠ جنيه .

٦- معامل الارتباط باستخدام النتائج (٢) ، (٣) .

٧- معامل الارتباط بطريقة الرتب .

٨- قارن بين النتائج المتحصل عليها فى (١) ، (٦) ، (٧)

الحل :

لايجاد المطلوب لابد أولا من تكوين الجدول التالى :

الدخل (س)	الانفاق (ص)	ح س = (س − ٦٤)	ح س = (ص − ٥٠)	ح' س = $\frac{س - ٦٤}{٤}$	ح' س = $\frac{ص - ٥٠}{٢}$	ح'س ح'ص	ح'٢ س	ح'٢ ص
٦٤	٥٢	صفر	٢	صفر	١	صفر	صفر	١
٥٢	٤٠	−١٢	−١٠	−٣	−٥	١٥	٩	٢٥
٨٤	٦٠	٢٠	١٠	٥	٥	٢٥	٢٥	٢٥
٦٤	٥٢	صفر	٢	صفر	١	صفر	صفر	١
٧٦	٦٠	١٢	١٠	٣	٥	١٥	٩	٢٥
٥٦	٤٢	−٨	−٨	−٢	−٤	٨	٤	١٦
٦٨	٥٠	٤	صفر	١	صفر	صفر	١	صفر
٦٤	٥٢	صفر	٢	صفر	١	صفر	صفر	١
−	−	−	−	٤	٤	٦٣	٤٨	٩٤

• نلاحظ أن العمودان (١) ، (٢) يمثلان القراءات الاصلية س ، ص أما العمودان (٥) ، (٦) يمثلان الفروق المعدلة ح' س ، ح' ص

* من الجدول السابق نجد أن :

$$\bar{\text{س}}^{*} = \frac{\text{مجـ } \acute{\text{ح}}\text{ س}}{\text{ن}} = \frac{٤}{٨} = \frac{١}{٢}$$

$$\bar{\text{ص}}^{*} = \frac{\text{مجـ } \acute{\text{ح}}\text{ ص}}{\text{ن}} = \frac{٤}{٨} = \frac{١}{٢}$$

$$\text{ع}^{*٢}_{\text{س}} = \frac{\text{مجـ } \acute{\text{ح}}^{٢}\text{ س}}{\text{ن}} - \left(\frac{\text{مجـ } \acute{\text{ح}}\text{ س}}{\text{ن}}\right)^{٢}$$

$$= \frac{٤٨}{٨} - \left(\frac{١}{٢}\right)^{٢} = \frac{٤٦}{٨}$$

$$\text{ع}^{*٢}_{\text{ص}} = \frac{\text{مجـ } \acute{\text{ح}}^{٢}\text{ ص}}{\text{ن}} - \left(\frac{\text{مجـ } \acute{\text{ح}}\text{ ص}}{\text{ن}}\right)^{٢}$$

$$= \frac{٩٤}{٨} - \left(\frac{١}{٢}\right)^{٢} = \frac{٩٢}{٨}$$

٤ – معامل الارتباط بيرسون :

يلاحظ أن معامل الارتباط بين س ، ص هو نفسه معامل الارتباط بين $\acute{\text{ح}}$ س ، $\acute{\text{ح}}$ ص .

$$\therefore \text{ر} = \frac{\frac{١}{\text{ن}} \text{ مجـ } \acute{\text{ح}}\text{ س } \acute{\text{ح}}\text{ ص} - \bar{\text{س}}^{*} \bar{\text{ص}}^{*}}{\text{ع}^{*}_{\text{س}} \text{ ع}^{*}_{\text{ص}}}$$

$$= \frac{\frac{٦٣}{٨} - \frac{١}{٢} \times \frac{١}{٢}}{\sqrt{\frac{٤٦}{٨}}\ \sqrt{\frac{٩٢}{٨}}} = \frac{٦١}{\sqrt{٤٦}\ \sqrt{٩٢}} = ٠,٩٤$$

أى أن الارتباط بين الدخل والانفاق طردى قوى .

٢– خط انحدار الانفاق ص على الدخل س :

طبقا لطريقة الفروق المبدلة يمكننا ايجاد معادلة انحدار (ص / س) على مرحلتين على النحو التالى :

(أ) نرجد اولا خط انحدار ح َ ص على ح َ س

$\text{ح}'_{\text{ص}} = \hat{\text{أ}}\ \text{ح}'_{\text{س}} + \hat{\text{ب}}$

حيث :

$$\therefore \hat{\text{أ}} = \frac{\frac{١}{\text{ن}} \sum \text{ح}'_{\text{س}}\ \text{ح}'_{\text{ص}} - \bar{\text{ص}}^{*}\ \bar{\text{س}}^{*}}{\text{ع}^{*٢}_{\text{س}}}$$

$$= \frac{\frac{٦٣}{٨} - \frac{١}{٢} \times \frac{١}{٢}}{\frac{٦٤}{٨}} = \frac{٦١}{٤٦} = ١,٣٣$$

$$\hat{\text{ب}} = \bar{\text{ص}}^{*} - \hat{\text{أ}}\ \bar{\text{س}}^{*}$$

$$\frac{\ }{٢} - ١,٣٣ \times \frac{١}{٢} = -٠,١٦$$

∴ ح/ ص = ١,٣٣ ح/ س − ٠,١٦

(ب) ∴ خط انحدار ص على س هو :

$$\frac{\hat{ص} - ٥٠}{٤} = ١,٣٣ \left(\frac{س - ١٤}{٢} \right) - ٠,١٦$$

∴ $\hat{ص}$ = ٠,٦٦ س + ٧,١٢

٣– وبالمثل يمكن ايجاد خط انحدار س على ص وتكون معادلة الانحدار هى :

$\hat{س}$ = ١,٣٢ ص − ١,٣٢

٤– انفاق الاسرة التى يبلغ دخلها ٧٠٠ جنيه :

$\hat{ص}$ = ٠,٦٦ (٧٠٠) + ٧,١٢ = ٤٦٩,١٢ جنيه

٥– دخل الاسرة التى تنفق ٤٨٠ جنيه :

$\hat{س}$ = ١,٣٢ (٤٨٠) − ١,٣٢ = ٦٢٠ جنيه

٦– معامل الارتباط باستخدام معاملى الانحدار (باستخدام نتائج (٢) ، (٣))

$$ر = \sqrt{\hat{أ} \times \hat{أ}١}$$

$$= \sqrt{\frac{٦١}{٤٦} \times \frac{٦١}{٩٢}} = ٠,٩٤$$

مما يدل على وجود علاقة طردية قوية بين الدخل والانفاق .

٧- معامل الارتباط بطريقة الرتب :

لايجاد معامل الارتباط بطريقة الرتب نكون الجدول التالى :

س	ص	رتب س	رتب ص	فرق الرتب (ف)	(ف)٢
٦٤	٥٢	٤	٥	١-	١
٥٢	٤٠	١	١	صفر	صفر
٨٤	٦٠	٨	٧,٥	٠,٥	٠,٢٥
٦٤	٥٢	٤	٥	١-	١
٧٦	٦٠	٧	٧,٥	٠,٥	٠,٢٥
٥٦	٤٢	٢	٢	صفر	صفر
٦٨	٥٠	٦	٣	٣	٩
٦٤	٥٢	٤	٥	١-	١
					مجـ ف٢ = ١٢,٥

∴ معامل الارتباط بطريقة الرتب هو :

$$ر = ١ - \frac{٦ \text{ مجـ } ف^{٢}}{ن (ن^{٢} - ١)}$$

$$= ١ - \frac{٦ \times ١٢,٢}{٨ (٦٤ - ١)} = ٠,٨٥$$

٨- المقارنة بين النتائج (١) ، (٦) ، (٧)

لوحظ أن قيمة معامل الارتباط بطريقة بيرسون [التى حصلنا عليها فى (١)] هى نفسها الجذر التربيعى لحاصل ضرب معاملى خطى الانحدار [التى حصلنا عليها فى (٦)] ، ولكنها تختلف عن القيمة التى حصلنا عليها بطريقة الرتب [المطلوب فى (٧)] كما كنا نتوقع حيث استبدلنا القراءات الاصلية برتبها .

(٧-٢) كيفية استخدام الانحدار الخطى البسيط فى تعيين الاتجاه العام للسلاسل الزمنية :

السلسلة الزمنية هى قراءات أو مشاهدات احصائية تسجل عن متغير ما عند فترات زمنية متعاقبة . فدرجات الحرارة اليومية المسجلة فى مدينة القاهرة وعلى مدار فترة زمنية مناسبة (سنة مثلا) ، وعدد الحرائق او الوفيات الشهرية المسجلة فى منطقة ما (محافظة الشرقية مثلا) عن فترة زمنية [من ١ / ١ / ١٩٩٢ الى ٣١ / ١٢ / ١٩٩٥] وقيمة المبيعات الربع سنوية من سلعة ما (الخضروات المثلجة من انتاج شركة قها) والمسجلة فى سجلات الشركة عن الفترة من ١ / ١ / ١٩٨٠ الى ٣١ / ١٢ / ١٩٨٩) ، وعدد حوادث المرور

المسجلة نصف سنويا فى منطقة القاهرة الكبرى عن السنوات ١٩٩٠ حتى ١٩٩٥ مثلا ، كلها تمثل سلاسل زمنية تختلف فيها المتغيرات المقيسة والاوقات التى تسجل عندها المشاهدات ولكنها ومثيلاتها تمثل سلاسل زمنية وفقا للتعريف المتفق عليه للسلسلة الزمنية . مما سبق يلاحظ مدى الاختلاف بين العينة العشوائية البسيطة والسلاسل الزمنية حيث تظهر الاخيرة (المشاهدات) فى السلسة الزمنية غير مستقلة عن الزمن .

فاذا رمزنا للزمن بالرمز س حيث س١ ، س٢ ، ٠٠٠٠ ، س ن متغيرات تشير الى الزمن فى تعاقبة (أيام - أسابيع - شهور - اجزاء من السنة - سنوات كاملة) والى المتغير المقيس بالرمز ص حيث ص١ ، ص٢ ، ٠٠٠ ص ن تعبر عن قيم المتغير التى تسجل عند الفترات المتعاقبة من الزمن ، فانه يمكن تمثيل السلسلة الزمنية بيانيا بشكل خريطة الخط البيانى *Line graph* أو السلسة الزمنية *Historigram* ومن الطبيعى أن تتعرض قيم المتغيرات المسجلة على مدار فترة زمنية مناسبة لأثر عنصر أو أكثر من العناصر الاربعة التالية

١ - الاتجاه العام *Secular Trend*

٢ - التغيرات الموسمية *Seasonal Variation*

٣ - التغيرات الدورية *Cyclical Varuation*

٤ - التغيرات العشوائية *Irregular Variation*

وتقوم التفرقة بين هذه العناصر الاربعة على أساس درجة الانتظام وطرز الاثارة والعناصر الثلاثة الأولى تعكس انتظاما فى التغيرات التى تعرضها المشاهدات ، هذه التغيرات التى تنشأ عن مؤثرات يمكن تحديدها ، وقياس أثرها ودرجة انتظامها واتجاه مسار تلك التغيرات . أما العنصر الاخير وكما يتضح من تسميته فهو يعكس أثر عوامل عارضة لا يسهل تحديدها .

وإذا أشرنا الى المتغير بالرمز (ص) والى أثر الاتجاه العام بالرمز (ت) والتغيرات الموسمية بالرمز (م) والتغيرات الدورية بالرمز (د) والى التغيرات العشوائية بالرمز (ع) .

فانه يمكن التعبير عن أثر كلا من العناصر الاربعة على السلسة الزمنية بأحد نموذجين :

(١) النموذج التجميعى : وصيغته الرياضية هى :

$$ص = ت + م + د + ع$$

(٢) النموذج الضربى (نموذج حاصل الضرب) وصيغته الرياضية هى:

$$ص = ت \times م \times د \times ع$$

وأيا كان النموذج المستخدم فى التعبير عن العناصر الاربعة للسلسله الزمنية فان الهدف من تحليل السلاسل الزمنية عموما هو :

١- التنبؤ بقيم المتغير (ص) مستقبلا وبالتالى التخطيط الجيد لمواجهة أى تغيرات متوقعة فى (ص) .

٢- تحديد وقياس أثر كل من العناصر الاربعة على قيم (ص) مما يساعد فى تخليص قيم المتغير (ص) من أثر كل منها ومن ثم المساعدة فى تعيين أثر بقية المؤثرات الاخرى بصورة أفضل مما يساعد فى اتخاذ القرار المناسب بشأن(ص) .

فى الجزء التالى سوف تقتصر مناقشتنا على تعيين الاتجاه العام حسابياً بطريقة المربعات الصغرى العادية (التى سبق أستعراضها فى بداية هذا الباب) .

تعيين الاتجاه العام بطريقة المربعات الصغرى .

سبق أن ذكرنا فى بداية هذا الباب أن معادلة الانحدار عموما تأخذ الصورة :

$$\hat{ص} = \hat{أ} س + \hat{ب}$$

وطبقا لطريقة المربعات الصغرى فقد أمكن ايجاد حلول لقيم الثوابت $\hat{أ}$ ، $\hat{ب}$ باستخدام المعادلتين الطبيعيتين كما ذكرنا من قبل . الا أنه فى حالة السلاسل الزمنية فانه يمكن تبسيط العمليات الحسابية بشكل واضح اذا نقلنا نقطة الاصل للمتغير المستقل (س) الى قيمة تتوسط قيم (س) بحيث يكون (مجـ س = صفر)

ويمكن تحقيق ذلك اذا كان عدد قيم (س) فرديا وذلك بنقل نقطة الاصل الى القيمة ذات الترتيب [(ن + ١) ÷ ٢] أى تكون نقطة الاصل هى السنة الخامسة مثلا وذلك اذا كانت ن = ٩ سنوات مثلا .

وبذلك تكون قيم س الجديدة هى :

-٤ ، -٣ ، -٢ ، -١ ، صفر ، ١ ، ٢ ، ٣ ، ٤

أما اذا كانت (ن) عددا زوجيا من السنوات فيتحقق ذلك بنقل نقطة الاصل الى منتصف الفترة بين السنتين الوسيطتين ، فمثلا اذا كانت ن = ١٠ سنوات ومن ثم تم أخذ نقطة الاصل فى منتصف السنة الخامسة وتكون قيم س هى

$$\frac{-٩}{٢} ، \frac{-٧}{٢} ، \frac{-٥}{٢} ، \frac{-٣}{٢} ، \frac{-١}{٢} ، \frac{١}{٢} ، \frac{٣}{٢} ، \frac{٥}{٢} ، \frac{٧}{٢} ، \frac{٩}{٢}$$

واذا تم ضرب قيم (س) فى ٢ فتكون قيم س المعدلة هى :

-٩ ، -٧ ، -٥ ، -٣ ، -١ ، ١ ، ٣ ، ٥ ، ٧ ، ٩

وبذلك تكون تقديرات أ ، ب سواء لعدد فردى أو زوجى من السنوات هى :

$$\hat{أ} = \frac{مجـ \ س \ ص}{مجـ \ س^{٢}} \quad ، \quad \hat{ب} = \frac{مجـ \ ص}{ن} \quad حيث \ [\ مجـ \ س = صفر \ ]$$

مثال (٥) :

الاتى بيان بالمبيعات السنوية لإحدى السلع الاستهلاكية فى مدينة الزقازيق

(بالمليون جنيه) :

السنة	١٩٧٠	١٩٧١	١٩٧٢	١٩٧٣	١٩٧٤	١٩٧٥	١٩٧٦	١٩٧٧	١٩٧٨
المبيعات	٧٦٤	٧٥٨	٨١٢	٨٢٩	٨٧٢	٩٨٠	١٠٣٩	١٠٠٦	١١٤٠

المطلوب : تعيين معادله الاتجاه العام لمبيعات هذه السلعة . والتنبؤ بمبيعات هذه السلعة خلال عامى ١٩٧٩ ، ١٩٨٠ .

الحل :

لايجاد معادله الاتجاه العام لمبيعات السلعة خلال الفترة ١٩٧٠ - ١٩٧٨

نكون الجدول التالى :

السنه	المبيعات (ص)	الزمن (س)	س ص	س٢
١٩٧٠	٧٦٤	-٤	-٣٠٥٦	١٦
١٩٧١	٧٥٨	-٣	-٢٢٧٤	٩
١٩٧٢	٨١٢	-٢	-١٦٢٤	٤
١٩٧٣	٨٢٩	-١	-٨٣٩	١
١٩٧٤	٨٧٢	صفر	صفر	صفر
١٩٧٥	٩٨٠	١	٩٨٠	١
١٩٧٦	١٠٣٩	٢	٢٠٧٨	٤
١٩٧٧	١٠٠٦	٣	٣٠١٨	٩
١٩٧٨	١١٤٠	٤	٤٥٦٠	١٦
المجموع	٨٢٠٠	صفر	٢٨٥٣	٦٠

ويمكن تقدير قيم الثوابت $\hat{أ}$ ، $\hat{ب}$ على النحو التالى :

$$\hat{أ} = \frac{\text{مجـ س ص}}{\text{مجـ س}^{٢}} = \frac{٢٨٥٣}{٦٠} = ٤٧,٦$$

$$\hat{ب} = \frac{\text{مجـ ص}}{\text{ن}} = \frac{٨٢٠٠}{٩} = ٩١١,١$$

وبالتالى يمكن كتابة معادلة الاتجاه العام على الصوره :

$$\hat{ص} = ٤٧,٦ \text{ س} + ٩١١,١$$

وعاده ما تعرف معادلة الاتجاه العام بتحديد نقطة الاصل بالنسبة لمتغير الزمن (س) وكذلك وحدات القياس لكل من المتغيرين س ، ص حتى لا يحدث اى خلط . وعليه فتكتب معادلة الاتجاه العام على النحو التالى :

$$\hat{ص} = ٤٧,٦ \text{ س} + ٩١١,١$$

(نقطة الاصل : ١٩٧٤ ، ووحده الزمن س : سنة ، ص : جملـة قيمـة المبيعات السنوية بالمليون جنيه).

∴ اجمالى قيمة مبيعات الشركة سنة ١٩٧٩ بفرض ثبات نفس الظروف السائدة هو :

$$\hat{ص}_{(\text{س} = ١٩٧٩)} = \hat{ص}_{(\text{س} = ٥)} = ٤٧,٦ \times ٥ + ٩١١,١ = ١١٤٩,١ \text{ مليون جنيه}$$

اما اجمالى قيمة مبيعات الشركة سنة ١٩٨٠ بفرض ثبات نفس الظروف السائدة هو :

$\hat{ص}_{(س = ١٩٨٠)} = \hat{ص}_{(س = ٦)} = ٤٧,٦ \times ٦ + ٩١١,١ = ١١٩٦,٧$ مليون جنيه

مثال (٦) :

الجدول التالى يوضح قيمة المنفق سنويا على التجهيزات والانشاءات فى بعض الصناعات بالمليون جنيه فى الفترة من سنة ١٩٨٥ الى ١٩٩٤ .

السنة	١٩٨٥	١٩٨٦	١٩٨٧	١٩٨٨	١٩٨٩	١٩٩٠	١٩٩١	١٩٩٢	١٩٩٣	١٩٩٤
قيمة المنفق	٢٢	٢٧	٢٩	٢٨	٣٢	٣٢	٣٠	٣١	٣٨	٤٦

المطلوب :

ايجاد معادلة خط الاتجاه العام لهذه السلسلة الزمنية

الحل :

لتعيين معادلة خط الاتجاه العام للسلسلة الزمنية لاجمالى المنفق على التجهيزات والانشاءات فى بعض الصناعات ننشأ الجدول التالى :

السنة	اجمالى المنفق (ص)	الزمن (س) مقاس من نقطة الأصل (١٩٨٩ - ١٩٩٠)	س٢	س ص
١٩٨٥	٢٢	٩-	٨١	١٩٨-
١٩٨٦	٢٧	٧-	٤٩	١٨٩-
١٩٨٧	٢٩	٥-	٢٥	١٤٥-
١٩٨٨	٢٨	٣-	٩	٨٤-
١٩٨٩	٣٢	١-	١	٣٢-
١٩٩٠	٣٢	١	١	٣٢
١٩٩١	٣٩	٣	٩	٩٠
١٩٩٢	٣١	٥	٢٥	١٥٥
١٩٩٣	٣٨	٧	٤٩	٢٦٦
١٩٩٤	٤٦	٩	٨١	٤١٤
المجموع	٣١٥	صفر	٣٣٠	٣٠٩

١٩٩٠) أى ١٩٨٩/٧/١ وبالتلى فإن مقدارات المربعات الصغرى لثوابت المعادله هى :

$$\hat{أ} = \frac{مجـ\ س\ ص}{مجـ\ س^{٢}} = \frac{٣٠٩}{٣٣٠} = ٠,٩٣٦ \quad ،$$

$$\hat{ب} = \bar{ص} = \frac{٣١٥}{١٠} = ٣١,٥$$

وتكون معادلة خط الاتجاه العام :

$$\hat{ص} = ٠,٩٣٦\ س + ٣١,٥$$

(نقطة الاصل : ١٩٨٩ - ١٩٩٠ ، س : وحده الزمن بالسنه ، ص : اجمالى المنفق على التجهيزات سنويا بالمليون جنيه) .

أخيرا ، ارجو التنوية إلى أن الأبواب السبعة السابقة ما هى الا محاولة متواضعه من المؤلف لما يمكن ان يقال وليس كل ما يقال فى هذه الموضوعات متمنياً ان تكون قد القت ولو بصيص من الضوء للقارىء المهتم بهذا العلم الحيوى .

المؤلف ،

المراجع العربية

- إبراهيم موسى عبد الفتاح (١٩٩٥) ، مقدمه فى الاحصاء الوصفى ، مكتبة المدينة بالزقازيق .
- أحمد عباده سرحان (١٩٧١) ، مقدمه فى طرق التحليل الاحصائى ، دار الكتب الجامعيه – القاهرة .
- جلال الصياد وأخرون ، مبادئ الاحصاء لطلاب الدراسات الاقتصادية والادارية جامعة الملك عبد العزيز – جده .
- حسن محمد على وأخرون (١٩٩٦) ، مقدمه فى الاحصاء التطبيقى ، مكتبة المدينة بالزقازيق .
- سمير كامل عاشور ، سامية أبو الفتوح سالم (١٩٩٠) ، مقدمه فى الاحصاء الوصفى ، معهد الدراسات والبحوث الاحصائية جامعة القاهرة .
- عبد اللطيف عبد الفتاح أبوالعلا (١٩٩٣) ، مدخل إلى الطرق الاحصائية ، مكتبة كلية التجارة دمياط الجديدة – جامعة المنصورة .
- عثمان على شلبى (١٩٩٥) ، الاحصاء الوصفى ، مكتبة المدينة بالزقازيق.
- محمد جمعه الروبى (١٩٩١) ، مقدمه مبادئ الاحصاء الوصفى ، مكتبة معهد الكفاية الانتاجية – جامعة الزقازيق .
- محمد عبد السميع عنانى ([illegible]) ، مبادئ الاحصاء الوصفى ، المكتبة العلمية بالزقازيق .

- Jonathan , D.C. , and Miller, R. B. (1991) , Statistics for Business Data Analysis and Modeling , PWS - Kent Publishing Company .

- Maddala, G. S. , Rao, C. R. , Vinod, D, (1993), Econometrics , North Holland .

- Walpole, R. E., (1982), Introduction to Statistics , Macmillan Publishing Company . Inc. , New York .

رقم الإيداع
٩٧ / ١٥٨٨
٩٧٧ - ٥٥٣٧ - ١٩
I.S.B.N.